AGARD-CP-458

AGARD-CP-458

AD-A212 884

# AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS No.458

# Human Behaviour in High Stress Situations in Aerospace Operations



NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

AGARD-CP-458

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Conference Proceedings No.458

# HUMAN BEHAVIOUR IN HIGH STRESS SITUATIONS IN AEROSPACE OPERATIONS

Papers presented at the Aerospace Medical Panel Symposium held in
The Hague, The Netherlands, from 24—28 October 1988.

## THE MISSION OF AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

— Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;

— Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);

— Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

— Improving the co-operation among member nations in aerospace research and development;

— Exchange of scientific and technical information;

— Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

— Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published June 1989



ISBN 92-835-0517-4

*Printed by Specialised Printing Services Limited*
*40 Chigwell Lane, Loughton, Essex IG10 3TZ*

## PREFACE

The operational performance of advanced military aircraft can be limited by human as well as aircraft constraints. In order to cope with this problem, rigorous medical selection standards have been introduced to eliminate medical causes of impaired aircrew performance. However, comparably rigorous psychological selection and monitoring procedures have not been implemented and employed. Moreover, the current techniques of psychological testing vary widely among NATO member countries. Experience has shown that many incidents and accidents may be attributed to resource limitations of the human operator even during normal air operations. In emergency and other high stress situations the demand on the operator may become overwhelming due to an increased and continous flow of information to be perceived, processed and acted upon. In fact, performance may deteriorate just at the moment when improvement is needed. Although certain aspects of this subject have been addressed in previous AMP Symposia, it was timely to look at the psychological limitations of aircrew behaviour more extensively.

This Symposium, which included papers from two non-NATO nations, dealt with the topic area t.om the point of view of incident and accident investigation, personality traits, responses to stress and prediction of behavioural responses. Although no specific conclusions were drawn, the presentations and resulting discussions were considered useful for aerospace medical specialists working in this field.

* * *

Les performances opérationnelles des aéronefs militaires évolués peuvent être limitées aussi bien par des contraintes humaines que par des considérations techniques. Afin de résoudre ce problème, des normes médicales de sélection rigoureuses sont appliquées afin d'éliminer les causes médicales d'une éventuelle dégradation des performances des équipages. Malheureusement, des procédures de sélection et de suivi psychologiques d'une pareille rigueur ne sont pas employées de façon systématique. En outre, les méthodes adoptées pour les épreuves psychologiques varient considérablement entre les différents pays de l'OTAN. L'expérience montre que bon nombre d'incidents et d'accidents peuvent être attribués aux limitations des moyens de l'opérateur humain, et ceci même au cours des opérations normales. En cas d'urgence et dans d'autres situations très critiques, l'opérateur risque d'être saturé par un flot accéléré d'informations qui doivent être perçues et traitées afin de lui permettre d'agir en conséquence. Les performances du pilote risquent donc de se dégrader juste au moment où un effort supplémentaire lui est demandé. Bien que certains aspects de ce sujet aient été traités lors de précédents Symposia du Panel de Médecine Aérospatiale, il a été considéré opportun d'examiner les limitations psychologiques du comportement des équipages de manière plus approfondie.

Ce Symposium, qui comportait deux présentations données par des conférenciers originaires de pays non membres de l'OTAN, a abordé le sujet du point de vue des enquêtes *sur les incidents* et les accidents, des traits de personnalité, des réponses au stress et de la prévision des réponses comportementales. Bien qu'aucune conclusion spécifique n'en ait été tirée, les présentations et les discussions qui ont suivi furent considérées comme étant d'un grand intérêt pour les spécialistes en médecine aérospatiale travaillant dans le domaine.

## AEROSPACE MEDICAL PANEL

Chairman: Colonel K.Jessen
Director of Aeromedical Services
Danish Defence Command
P.O. Box 202
DK-2950 Vedbaek
Denmark

Deputy Chairman: Mr Charles Bates, Jr
Director, Human Engineering Division
AAMRL/HE
Wright-Patterson AFB, Ohio
45433-6573
USA

## TECHNICAL PROGRAMME COMMITTEE

Prof. Dr. H.T.Andersen
Director
RNoAF Institute of Aviation Medicine
P.O. Box 14, Blindern
0313 Oslo 3
Norway

Dr. A.J.Benson
RAF Institute of Aviation Medicine
Farnborough, Hants GU14 6SZ
United Kingdom

Capt. W.M.Houk, MC USN
Commanding Officer
Naval Medical R&D Command
NMC/NCR
Bethesda, MD 20814-5044
USA

## HOST NATION COORDINATOR

LtCdr E.J.H.Bleeker
(RNLN Ret)
c/o National Aerospace Lab. (NLR)
Netherlands Delegation to AGARD
P.O.Box No.126
2600 AC Delft
The Netherlands

## PANEL EXECUTIVE

Major J.A.Winship, CAF
AGARD/NATO
7, Rue Ancelle
92200 Neuilly sur Seine
France

CONTENTS



* Not available at time of printing

*Not available at time of printing

TECHNICAL EVALUATION REPORT

The symposium "Human Behaviour in Crises Situations in Aerospace Operations" was sponsored jointly by the AMP "Human Factor" and "Special Problems in Physiology and Medicine" sub-panels. The planning went on for approximately one year and a half after the theme had been approved by the AMP.

It was realized early on that relevant data on human behaviour obtained during real emergency situations would be scarce since such incidences are unpredictable and therefore, an unlikely object of systematic research. For this reason the programme committee aimed at inviting research contributions obtained utilizing synthetic stress, for instance, experiments carried out in simulators and, likewise, papers based on psychological methods which would predict behaviour in high stress situations. Since selection, training and treatment of air crew may give additional information on the subject, authors with experience in these areas were also invited to contribute. In the end of the programme consisted of 24 papers - all of which but one were presented - divided among of four main topics, namely,

1. incident and accident investigation
2. personality traits of aviators, selection and training
3. air crew responses to stress
4. prediction of behaviour

The symposium started with a series of papers presenting the consequences of adverse events and their causes, methods of how to investigate incidents and accidents, and, to some extent, treatment of the air crew involved was discussed. In particular, past experiences of mishaps of the RAF and the RNoAF were presented together with the underlying psychological mechanisms. It is noteworthy that certain air forces outside the NATO community contributed to this part of the symposium. Thus, in addition to the national Air Forces already mentioned above, the corresponding records of the Swedish and the Israeli air forces were presented and discussed.

Next in the programme followed a session devoted to studies of personality traits of aviators.
The use and interpretation of the Defence Mechanism Test (DMT) were addressed by several speakers. Several conflicting views were submitted. In fact, the programme committee had specifically encouraged scientists whose opinions of the DMT differed to relate their various experiences because the test itself as well as its predictive value is controversial. Supporters of the DMT attached a great deal of significance to the results obtained provided the test was administrated by experienced personnel in order to obtain information which would not otherwise become revealed. Other workers maintained their sceptiscism as to the usefulness of the DMT.
Most Air Forces use psychological testing of candidates for military flying training, but whereas psychologists seem to have decisive influence on the selection process in certain nations, others attach much less significance to such advice. There exist presently a working group within the NATO community trying to establish a standard test battery which may be administered and evaluated in a reasonable period of time.

The third session dealt with air crew behaviour during stressful events. Obviously, the research presented had largely been carried out in simulators. It has been argued that human decision making during synthetic conditions is very different from approach taken in the real world. This idea fundamental to all simulator training was neither validated nor rejected at the symposium. It is difficult, however, to accept that simulator stress is not "real" to the extent that it would provoke fundamental behavioural mechanisms. However, it was pointed out in a discussion that the punishment of simulator faults is not severe enough. If penalties had serious consequences for the operator rigorous experiments would yield reproducible results!
Finally, some of the personal characteristics which might predict performance were discussed. One of the very interesting observations made was the wide variation of personal qualities focused upon. The Success of future fighter pilots was forecasted using such different scales as "time to first kill" in simulated air combat maneouvering - which was utilized as the standard in one paper - to how the priorities of the modern western society affect the general standards of adolescents - a point of view which was emphatically argued by another author, apparently without superficial resemblance whatsoever, these characteristics might be closely interrelated when looked upon in perspective.
In any event, the AGARD AMP had scientifically rewarding and very stimulating symposium.

# PREDICTION OF PERSONALITY

Harald T. Andersen M.D., Ph.D., D.Sc,D.Av.Med.
Director
RNoAF Institute of Aviation Medicine
P.O. Box 14, Blindern, 0313 OSLO 3, Norway

I would like initially to stress our primary objective as spesialists of aviation medicine and aviation psychology for different national air forces by recalling an episode which took place quite recently.
A firm believer in technological solutions to most of the problems encountered by piloted, airborne weapon platforms put this question to me: "in this time of incredible technological achievements what is the contribution of flight medicine to military air power?"
Admittedly a bit exuberant, I responded: "We provide the marginal edge of superiority!"
My argumentation obviously being that recent political developments aimed at reducing the potential for mass disaster through diminished capacity for nuclear warfare, and, a concurrent narrowing of the technological gap between opposing forces, cause increased emphasis to be placed on conventional military power. Thus, the side provided with the meticulously selected, better trained, confident and well prepared warriors will be at advantage, however slight.
The argument actually went one step further with an acrimonious second question being asked: "and how do you do it?"
The audacious answer on behalf of all of us had to be:
"we pick the winners and we train them!"

Coaching of modern military aviators is a game of producing winners against considerable odds, the likelyhood of being tossed against well trained and well equipped opponents is only one of the factors to be taken into account. Others are the limits to physiological and psychological capabilities of air crew, well known to flight surgeons and aviation psychologists, but largely neglected by designers of airborne weapon systems, and frequently minimized by operational leaders.
Spesialists in aviation medicine - psychologists not excluded - have contributed greatly to prevent public investments in unsuitable candidates by carefully choosing among applicants for military flying training by applying exact clinical principles, advanced methods of examination and statistical knowledge of the prognosis of disease in the general population to the selection procedures. Therefore, the possibility of organic disease to occur in air crew populations before the age of 40-45 years when a flying carreer i usually terminated by assignments outside the community of operational fliers, is a rather remote one. In fact, the future development of health profiles of the post-industrialized society may show itself as more of an obstacle to aeromedical selection boards than additional advances in examination techniques and diagnostic procedures.
These same arguments are not equally applicable to the behavioural sciences - for two different reasons. First, modern technology has been introduced into the cockpit environment to an extent which is not always helpful to the pilot. Quite to the contrary, one may state that the amount of information presented to the military aviator easily cause over-loading of sensory channels, perceptive analysis or data processing ability. We know that even experienced pilots when in distress reduce tension in order to concentrate on the flying task by choosing to ignore or even to switch off information which annoy in competing for sensory attention. Obviously, the operator has become saturated and have decided that much of the information presented is irrelevant to his primary objective. Moreover, we also know that due to careful medical screening and relative safety of technical features of aeroplanes 60-80 per cent of all incidents and accidents are referred to the human factor, not only in advanced fighters, but in fairly simple aircraft as well. It is appropriate, indeed, to have the AGARD AMP to look at the behaviour of air crew in crisis situations in some detail, such is the aim of this symposium.
Approximately one year ago I accepted to chair the programme committee for this symposium. During this past year I have worried a great deal about what the audience might expect from those of us who undertook the task of putting the agenda together. At times I have been apprehensive that the symposium would come to belie its title since data on human behaviour are not easily collected during high stress situations in aerospace operations. The best your programme committee has been able to offer in the way of systematic research is evidently,

1) to bring together workers who have spent considerable time and effort investigating incidents and accidents,

2) to obtain contributions from colleagues with experience in studying personality traits common among aviators,

3) to have experts present opinions on how air crew may respond differently to stress, and, finally

4) to show to what extent selection and training may predict behaviour.

A discussion of these key-notes may aptly start by quoting from a paper which Drs STATLER & BILLINGS presented to the NATO Military Committe and various other expert groups within the alliance earlier this year. Their paper entitled "Military Pilots Ergonomics" published in AGARD HIGHLIGHTS 88/2 contains an intriguing statement which takes us directly to the core of the problem which we are about to address:

"..... humans make decisions very differently in a stressful and rapidly-changing real world than they do the sterile non-threatening environment of the laboratory."

This is an exciting challenge to present research in psychology as applied to aviation. If the word "simulator" is substituted for "laboratory", it becomes a formidable threat to certain current concepts of teaching flying skills, not to speak of the many businesses which depend on making such technology available for training purposes.
The statement of STATLER & BILLINGS is ambiguous, of course, and needs to be examined more closely. The authors may possibly suggest that laboratory work does not provide us with the information we need in order to understand, thus, leaving human decision-making forever enigmatic. The other reasonable interpretation of the statement is that psychologists might be able to predict behaviour correctly, but that students of the behavioural sciences at the present time do not fully recognize the significance of their laboratory findings.
Obviously, if air crew arrive at decisions acting differently during stressful events depending on circumstances, the implication is not necessarily one of unpredictable behaviour. When, in fact, the scientific observer is surprised by a decision made by an over-loaded pilot, he ought to share his amazement with the flier. The aviator, admittedly, may be astounded to discover what he has done, whereas the observer might be astonished to have been caught unprepared not anticipating what would happen due to ignorance, muddled thinking and deficient scientific methods. Fortuneately for all of us STATLER & BILLINGS moves on immediately to clarify their expression adding

"What we do *not* understand is the process of cognition"

No-one, I am sure will have difficulty accepting this statement. However, I am not convinced that reactions to stress might be exactly foretold albeit helped by a better understanding of cognitive psychology, the reason being that cognition is only one of many containers in the Box of Pandora. What we really are in search for is a psychological entity by means of which behaviour would turn out predictable.
A firm relationship between personality and behaviour, has actually been suggested to exist very much like a law of nature by Raymond B. CATTELL in one of the famous statements of personality theory:

"Personality is that which permits a prediction of what a person will do in a given situation."

According to CATTELL, therefore, the behaviour of operators - even when unexpected to observers - would show itself predictable if scientists were capable of comprehensively exploring personality. One difficulty encountered by researchers devoted to unveiling fundamental principles of behaviour is the lack of fixed constants, numbers or characteristics which might be assumed not to change with time or circumstances.
Well defined, unchangeable concepts are central to the progress of any scientific discipline. Students of mathematics and the natural sciences are fortunate to know a number of constants which permit strong inference and rapid advances within their chosen fields of study. Mathematicians, for instance, may work with marvels like pi = 3.1416 - today determined with more than 10 million decimals, Euler's number e = 2.71828 ... and many marvellous trigonometric relationships. Physicists are equipped with the laws of thermodynamics, knowledge of elementary particles and quantum mechanics. Chemists are provided with the Periodic Law of Elements, the nature of chemical bonds and the properties of reactions. Biologists build their wisdom on the cell concept and are well aware of energy transformation through a multitude of metabolic pathways. Genetecists know the DNA molecule and even how to manipulate it.
Although the possession of exact references is to be envied by the rest of us, the behavioural sciences are not quite without rigorous methods although largely of a descriptive nature. Personality traits may be carefully measured qualitatively as well as quantitatively. Moreover, it has been strongly argued that anatomical structures, physiological functions and neurological considerations are connected in a causal way to the concept of personality, making H.J. EYSENCKs statement

"Personality is the fundamental unit of psychology"

quite plausible. Except for future possibilities of genetic manipulation, it may be inferred that the frames of reference of every characteristic feature of any individual is determined by the genetic material of the germ cells. Yet, the structure of the personality may appear to us almost as infinitely adaptable to stresses imposed by the environment through the process of learning. For practical purposes of aviation psychology, therefore, the crucial question is whether the methods for selection, training and treating air crew provide information which may be

interpreted accurately in order to predict behaviour along a scale of arousal ranging from complacency to panic.
Statements of the obvious are largely to be discouraged in audiences of well informed, knowledgable colleagues with a background in science or military leadership - or both. Nevertheless, I will risk submitting the observation that in military aviation writings on the wall, however evident to ardent observers, may not be appreciated, interpreted or acted upon by those for whom the information was intended. The demand for immediate recognition of unintentional deviations from planned strategies, which require proper corrective actions early on, is sometimes neglected by operators, enticing them into crisis situations. Professional intervention before or after such episodes may decisively influence human behaviour during high stress situations in aerospace operations.
Admittedly, flight surgeons and aviation psychologists are not usually present to perform studies or services during actual events - which is neither neccessary nor desirable. However, by selecting, training and treating those involved in operational tasks we prepare people for highly skilled, exceedingly demanding and, sometimes, dangerous tasks by being able to register how well or poorly they perform when severly stressed.
Selection and training are determinants of operative behaviour. Assuming that recruitment procedures have provided an attractive lot of candidates for military flying training there are still differences of opinion as to whom should be selected.
There are air forces represented here whose psychologists have a decisive influence in selecting pilots. In other air forces, to the contrary, the belief in psychological selection before initial flying training is not equally strong. Since behavioural scientists seem to hold conflicting views among themselves with respect to which tests are the most discriminative, and not everyone are convinced of the predictive potential of tests which others appear to rely heavily upon, we are looking very much forward to have these various opinions tossed at each other.
The most interesting shift of attitude presently taking place, however, is the apparent change of priorities as applied to personality traits among military aviators. Aggressivness seems to be less attractive than before. Safe operators are desirable, crew co-ordination is emphatically taught in mandatory courses. The question is whether these various demands may be selected for simultaneously providing us with personalities which may be trained for all of the desirable purposes. It is not unlikely that such a possibility exists. On the other hand, it is not probable that training commands will be very successful unless operational authorities have their minds made up to know exactly what should be selected for and trained. Because organizations like the AGARD-AMP may exert considerable influence over operational philosophies, it seems advantageous to look at our own priorities first.
The privilege of giving this address calls, I believe, for stating my own position on this question.
I believe that effective use of air power depend on aggressive fighter personalities. Controlled aggression is essential to any kind of tactical flying including transport, support not to forget search and rescue. Military aviation is a deadly serious competition against highly skilled opponents or extreme forces of nature and successful pilots are highly competitive. In my opinion we shall have to continue selecting for aggressive behaviour in stable individuals.
There are colleagues, I know, who believe that our present selection of candidates is done within populations less than ideal for other advantageous characteristics such as potential for system operation, safety through excellent abilities for crew coordination even outstanding academic performance. In my opinion, all of these remarkable features are highly desirable. However, they cannot substitute - merely serve as complements - to an indispensible, basic aggressiveness. Moreover, such personality traits neither prevent crisis situations to occur nor do they guarantee that tension and stress are handled well.
For instance, on take off with a DC 10 from a large well equipped international airport the cockpit crew, who had very recently been through a course in crew coordination given by the airline, exchanged polite niceties while the aircraft was rapidly approaching stall speed. No one had really assumed command and no one flew the airliner. They were brought back to the real world when much to their dismay, the flaps came out automatically. Obviously, it is possible to be overtrained for a good purpose. To take another example, crew coordination is not confined to a flight deck with a number of positions for different crew members. Everyone, who has even very limited experience with fighter flying will agree, I am sure, that flying a tight formation at low level or engaging in air combat maneouvering demand cooperation a bit beyond ordinary measures. The exercise constitute an extended period of very high stress which is - as a rule - dealt with very well, indeed.
I believe that at present selection procedures are such that basically aggressive but stable candidates have potential to be trained in disciplines which further safe operations. Moreover, I agree that our imperfect understanding of cognition and other psychological mechanisms makes life difficult and sometimes dangerous for military pilots. It is quite safe to state that correction of errors is a more frequent practice of air crew than is avoiding making mistakes altogether.
Since the time interval in which to respond is incredibly short, it follows that systematic recording of useful scientific data is impractical in the real world.
For this reason, I believe that laboratory experimentation is a source the yield of which will continue to throw additional light on human behaviour in crisis situations. The challenge is really on the part of the scientific imagination to define problems, conduct relevant experiments, obtain dependable results and to interprete the results correctly perhaps even to the extent of verying CATELLs statement:

"Personality is that which permits a prediction of what a person will do in a given situation."

Accepting the argument that the mechanisms by means of which stresses are resolved and coping is furthered are time dependent during military flying and other similarly skilled activities, it follow that human behaviour in crisis situations is a function of preparedness. Preparedness is the key to predictable behaviour, and, itself a function of psychological selection, training and treatment. In fact, I would like to conclude by stating that

PREDICTION OF BEHAVIOUR IS THE ESSENTIAL GOAL OF APPLIED AVIATION PSYCHOLOGY.

# 'HUMAN BEHAVIOUR IN HIGH STRESS SITUATIONS IN AEROSPACE OPERATIONS'
## KEYNOTE - THE OPERATOR'S PERSPECTIVE

by

Wing Commander Nicholas B. Spiller RAF
Chief, Fighter Branch
Defensive Operations Division
Headquarters
Allied Air Forces Central Europe
6792 Ramstein Air Base, Germany

SUMMARY

Air Power, by its nature is aggressive. Consequently, men with aggressive fighting qualities are needed to man front-line combat aircraft. The qualities that go to make a 'fighter' are not necessarily those required of safe peacetime pilots. As peace in Europe continues, and the cost of aircraft increases, there could be increasing pressure to man NATO's air forces with safe rather than aggressive pilots.

The modern aerial battlefield is becoming ever more complex. Modern air-to-air systems can give details of 20 or more targets at ranges of over 100 nms. Yet weapons, such as AIM-54 Phoenix are capable of engaging an aircraft from an area of 39,000 sq kms. Missile performance on this scale will force designers to provide even more information on the local air picture. The problems of data deluge are likely to increase rather than decrease.

The modern air combat arena is becoming more deadly and gladiatorial. In a future Northwest European conflict, traditional attrition rates could well double or triple. Aircraft and crews will almost certainly be required to fly several missions per day. The strain of intensive combat flying with high loss rates, combined with the stress of air attack while resting between missions and the physical fatigue of operating in NBC equipment, could place an intolerable psychological burden on the aircrew.

## THE PILOT's PERSPECTIVE

Gentlemen. It is a great pleasure and a great honour to be here addressing you today.

I suppose I should, at the outset, lay my cards on the table and admit my complete lack of credentials to be here. I am not a psychologist or a behavioural scientist.

My comments today are those of the user. The man who has to motivate and fly with the young men that you choose for our air forces. My lack of formal qualifications could be considered a disadvantage, but I believe that, for me at least, it offers one big advantage. I can happily spend the next 20 minutes flagging up problems and then at the end say "Thank you, Gentlemen. You are the experts. I leave the solutions to you."

Let me begin with a base plate on which to build my case.

Research during World War II and the Korean War - admittedly on infantrymen, but I think that much of it can be read across to the Air - showed that generally people could be split into two groups - the 'fighters' and the 'non-fighters.' By this, I mean that the 'fighters' are the ones who actively engage the enemy and the 'non-fighters' are the ones who, although they accept the same risks and dangers, have something in their psychological make-up that discourages them from taking an active part in the fight.

As we are dealing exclusively with Aerospace Operations at this Symposium, it is perhaps worth remembering that Air Power, by its very nature, has to be aggressive. In the Air Force, more than in any of the other Services, we rely on our young men to go out and actively engage the enemy.

When a Tornado crew takes off, they are expected to take their bombs deep into the enemy hinterland and while doing so, take on the full might of the enemy.

Even the air defence pilot on this side of the border has to be aggressive. His aim must be to actively seek combat, to go out looking for a fight.

'Fighters', fighting men with an aggressive spirit, therefore, are absolutely vital to us in the air power business if we are to make the proper use of our costly pieces of aeronautical hardware.

Because of this, let me spend a few moments outlining some of the qualities that the Korean War researchers said go to make a 'fighter'.

First is <u>leadership</u>. Under this heading, they include poise, spontaneity, extraversion, freedom from anxiety and independence.

Second is what the Americans might call the masculinity factor but which I, with British understatement, prefer to call outdoor adventurousness.

The third requirement is intelligence. The more intelligent a man, the more likely he is to be effective in combat.

Fourth, is a sense of humour. A sharp wit, preferably with biting sarcasm is what is needed. The ability simply to tell jokes is not enough.

Finally, emotional stability is vital. 'Fighters' were found to be more stable, less anxious and less prone to depression than their non-fighting colleagues.

*Turning for a moment to* 'non-fighters'. The research showed that 'non-fighters' tended to come from poorer homes. Less of their families owned their own businesses and generally they seemed to have a far smaller financial stake in their country than the 'fighters' did. In fact, their family life in all its aspects appeared to be more unstable and less harmonious than the 'fighters'.

Bringing all this together, we end up with a list of qualities for our ideal fighter pilot looking something like this:

Leadership
Masculine/Adventuresome
Intelligence
Stable Extroversion
Patriotism.

The first thing to notice about this list is that, with the arguable exceptions of Patriotism and Leadership, the qualities that we require in our fighter pilots are largely in-built. You either have them or you don't. They can't be instilled after the man has joined the Service.

In other words, we have to make sure that we recruit people with the right characteristics in the first place.

A glance round the average Squadron crewroom would probably convince most of you that we do, in the main, recruit fighter types.

A study by the United States Navy of its jet pilots showed that they were high on exhibitionism, dominance, heterosexuality and aggression and low on deference, affiliation and abasement. I doubt that their Air Force counterparts are very dissimilar.

So what is the problem? We need fighters to fill our cockpits and we have them.

My concern is this; extroverts - particularly exhibitionist, dominant, heterosexual, aggressive extroverts - do not make safe, peacetime pilots.

Having said that, I do have to admit that all is not bad. In some ways extroverts do outperform introverts - but they also offer the biggest risk to peacetime flight safety because their impulsive and sensation seeking nature tends to lead them into situations which exceed the limits of their flying skills.

Emotive words? Away from the microphone I might admit to a few hair-raising incidents of my own. I could certainly list a large number of aircraft accidents, from all nations, that fall into this category.

In the past, the aircraft lost in these accidents were generally aircraft like the Phantom and the F-104 costing no more than 3 or 4 million dollars.

We are now in the era of aircraft like the 20 million dollar Tornado. Harrier GR5 is quoted to cost rather more and I have no doubt that EFA, ACA and ATF will provide another quantum leap in expense.

The loss of aircraft costing that sort of money in peacetime is not going to rest easy with our governments. Particularly when defence budgets are contracting in real terms and there is no corresponding reduction in commitments.

As defence equipment becomes more costly and so more scarce, there will be increasing pressure to reduce our accident rates.

Aircraft today are very reliable. As you gentlemen know only too well, more often than not, it is the people in the loop that cause the problems.

When we respond to the accountants' call to reduce accident rates we must be very sure that we do not end up recruiting people for the wrong reasons. Safe, peacetime pilots sitting in the cockpits of our tactical fighters may be fine in peacetime, but could we be sure that they would go out and actively engage the enemy in wartime. If not, we may as well not bother to have them in the first place.

Let me turn to something different. I want to talk for a moment about the environment that we expect our future fighter pilot to fight in.

In World War II, the pilot of a fighter aircraft had a weapons system that had an effective range of no more than 500 metres. His sensor system, his eyes, had a range of several miles. It was, if you like, a matched system. He could see anything long before he could engage it and, in theory at least, he could see any potential attackers before they brought their guns to bear. To a large extent, the same was true during the Korean War.

It was the advent of the air-to-air missile that changed the rules. Suddenly, we had a weapon that could go as far, or even further, than the pilot could see. To produce a matched system we had to increase the pilot's detection range to maximise the capabilities to the weapon.

The result of this logic is epitomised today in aircraft like the F-15. These aircraft have a radar that can look in excess of 100 nms, can track several targets accurately or, in a raid assessment mode, give information on many more targets. The weapons include semi-active, radar missiles, all-aspect, infra-red missiles and cannon. For self defence, there are chaff and flare dispensers.

Aircraft of this calibre make great demands on the pilot. Not only does he have to be aware of what is going on around him visually, but he must also make some sense of the 20 or so blips on his radar tube, know the capabilities, limitations and engagement criteria of his three different weapon systems and all the while fly and fight the aeroplane.

To make matters worse, the time available for him to sort out all this data has been dramatically reduced.

In World War I, two state of the art fighter aircraft heading towards each other would close at roughly four miles/minute. By World War II, the closure rate had risen to 10 miles/minute. By Korea it was 17 miles/minute and today it is about 38 miles/minute.

It would be nice to say that these problems of data deluge - for that is what we are talking about - are going to go away. To say that modern electronics will provide an answer and solve the problem. I fear that the opposite will be the case.

Modern engineers are justifiably proud of the latest airborne radars that can track-while-scan 20 targets - more than enough for the average brain to cope with. But let us look at the other side of the coin.

The Phoenix AIM-54 air-to-air missile has already recorded hits during tests at ranges of over 100 nms. The Russian AA-9 is credited with similar performance. An aircraft would be vulnerable to a hit from missiles like these from an area of 39,000 sq kms. In a war in Northwest Europe, there are likely to be at least 100 aircraft in an area of that size, and during intensive operations there could be as many as 1000 aircraft. While those sort of imperatives are driving designers, they are not going to stop at a capability to track 20 aircraft. Modern electronics are going to be used to increase capability before they are ever used to ease the pilot's workload.

Another factor that needs to be considered about the modern battlefield is its lethality. In World War II, Emil Lang shot down 18 aircraft in one day, Marseille 17 aircraft in a day and Hans Hahn shot down 40 aircraft in seven missions. These scores were all with fixed forward-firing guns. Since then the guided air-to-air missile has made crackshots of those of us who might otherwise have missed.

AMRAAM will soon be coming off the production lines. Active missiles such as AMRAAM will make the air war even more deadly. Soon, aircraft equipped with active missiles will need to approach no nearer than 30 miles to their targets before launching their missiles and turning away. No loitering to keep the target illuminated - just fire and forget.

The trouble is that if both sides have that sort of capability, then there is every chance that both sides will launch effective missiles and that we will all be killed.

We aircrew may not be very smart but even we are likely to realise that those are not good odds for our survival.

There is a point here for our Air Forces. I know that my Air Force and, I think, all the Air Forces of Western Europe, have extended the lives of their front-line aircraft to make them last until the funds are available for a new model. It is a convenient way of keeping the front-line numbers up.

Our pilots are not Kamikaze by nature. They will willingly go on a mission as long as they consider that they have a reasonable chance of returning. They will not be prepared to go on what they believe is going to be a one-way mission. Because of this our pilots have to have faith in their aircraft - they have to believe that it will give them a reasonable chance of success in a future conflict.

If they don't have that faith in their equipment, they will either become disillusioned and leave - and we have seen a fair bit of that recently - or they will soldier on for their career or job security. And if they are doing that - can we be sure that they will do their duty when we need them to go out and fight in their 20-year-old warplanes?

Let me put another factor into this already complex equation of World War III. In the Central Region we will start the war about 2 to 1 down in tactical aircraft.

To offset the Warsaw Pact's numerical superiority, we will have to maximise our sortie rate. This will mean for our aircrew that the pace of World War III will be very much faster than we have been led to expect from history.

RAF Bomber Command between 1942 and 1945 often did not fly for weeks because of the weather. The US 8th Air Force had a similar pace of operations. Both forces overall sustained about 1½% attrition.

Some people talk about us sustaining 5% attrition in the next war and suggest that this would be acceptable. Between 1942 and 1945, the US 8th Air Force operated 10,500 aircraft, of which 5500 were destroyed by enemy action - over half their total fleet - and this with 1½% attrition. 5% is really going to hurt us.

In World War III our crews are going to have to fly 2 or 3 times a day. A force flying 3 sorties a day and suffering 3% attrition - not 5%, 3% - will suffer nearly 50% losses in a week.

For the statisticians playing their wargames that probably looks pretty reasonable. But how is it going to look to the pilots on the front-line? How is their morale and sense of duty going to look after a week of operations. A week where nearly half their comrades are dead or captured. A week of intensive operations with little sleep and probably hours wearing fatiguing NBC kit - of bombing attacks to break up their rest periods, confusion and probably some fratricide from their own forces.

It is going to take quite a leader to inspire those young men to go out and face the enemy again. And some very special young men to say "Yes, Sir" and go out and do it.

Somehow, from our democracies that have not known total war for 40 years, we have to recruit those kinds of people and, having recruited them, train them so that they will be prepared, mentally and physically, to survive the shock of the opening rounds of World War III.

I mentioned earlier AMRAAM and - the Beyond Visual Range battle, the BVR battle. There is no doubt that this battle is going to change the shape and flavour of Air Combat. While there is undoubtedly still a place for the old fighter pilot whose skill was in the close-in dogfight, in the future, before we can even think about the close-in engagement, we will have to think about, and win, the BVR battle.

This is going to do two things.

First, it is going to make the relative capabilities of the aircraft involved very much more important. If the opposition can out-see you and out-shoot you, then you start at a grave disadvantage. This brings me right back of course to the pilots having faith in their equipment. You can be Baron von Richthofen himself, but if the opposition can have a missile in the air before you even know that they are around, the chances are that you are going to lose.

The second point, and this is perhaps the more relevant to this forum, is that the BVR battle and the close-in dogfight are very different disciplines requiring very different skills.

The close-in combat is, if you like, a fairly straightforward affair, with logical moves and counter moves.

The BVR battle is more like a game of chess - it requires you to out-think the enemy. Height, speed, deception, stealth, and timing will all affect the outcome.

A person who excels in the close-in-dog-fight may well not show the same skill in the BVR battle. In fact, he may even lack the very attributes that he needs to excel in the BVR battle. It may be that we need a new kind of man for this new form of combat.

For once, I think that some of the solution to this problem may already exist. It seems ironic but it could be that our entertainment-orientated society is already starting to train the sorts of minds that we need.

I am talking, of course, about the computer revolution and all the computer games on the market that our youngsters are rushing out to buy. Some of them are quite impressive. There is one called ACES 2. We have a copy at home. I have not beaten my son at it for 2 years - he is 13 years old.

The Russians may be able to train their gymnasts from birth but, at the moment, technology and our consumer-orientated society is giving us an unconscious lead in the early training of our future fighter pilots.

I said at the beginning that I was not going to offer any solutions. Before I close though, I would like to give some personal thoughts on data deluge and pilot workloads. I fully appreciate that you are the experts.

The biggest message I have is that the aircraft itself has to be as easy to fly as possible. It is after all only a tool to get the job done.

I appreciate that that must sound like a blinding glimpse of the obvious.

On the other hand, I was amazed to see, the other day, on one of NATO's newest multi-million dollar fighters that, to know how much fuel there is, the pilot has to read a guage, press a switch to obtain a new reading and then add the 2 figures together. Surely we can do without that sort of hassle in a modern cockpit.

In the 1960s and 70s, I flew one of the most beautiful aircraft ever built. That aircraft had 2 big fuel guages high up in the cockpit, on the right-hand side. I forget the figures on those guages, but I do remember that when the needles were at 12 o'clock it was time to go home and that by 9 o'clock you had to be on your last circuit to land. We need the same sort of simplicity in our modern aircraft. Analogue guages can be interpreted at a glance. Digital readouts you have to think about - and that takes time.

That aircraft from the 1960s had some other nice features. To start it, you kicked one gangbar that put on all the switches, then pressed the engine starter button. It was so easy that you could do it while strapping in. With the capabilities that modern electronics give us, we should be able to have similar systems now in our front-line aircraft - one switch that switches on everything that you need to go flying. Subsequently, if something isn't working, or stops working, then the same system takes the necessary emergency actions and tells you what to do - to go home, to avoid negative G, or whatever.

We must use the capabilities of the micro-chip to thin out the aircrews' Flight Reference Cards. The present ones are far too thick to memorise and far too complex to be easily referred to in-flight.

Lastly, I think that there is still great scope for the imaginative use of colour in the cockpit. Operationally, an obvious use could include differentiating between friend, foe and neutral. It could also be used to show how current the information is - or how reliable the source. Airspeed indicators could change colour if the aircraft flies too slow or too fast. Altitude readouts could change colour as the aircraft nears the ground.

I have taken enough of your time. Thank you for your attention. Forgive me if I have taken you down paths that you know too well already, but I wished to show you some of the concerns from the operator's end of the telescope. I am, of course, willing to take your questions, but remember, you are the experts. I may not have the answer.

# CAUSES OF AIRCREW ERROR IN THE ROYAL AIR FORCE

J. W. Chappelow

R.A.F. Institute of Aviation Medicine
Farnborough
Hants. GU14 6SZ
U.K.

## SUMMARY

One hundred and forty nine military flying accidents were investigated by psychologists. Inspection of the data collected revealed that nearly half of the accidents involved inadequacies in equipment design, training or administration. Cognitive failure was a major cause of aircrew error and was more often associated with underarousal than with overarousal. Overarousal made a significant contribution to aircrew error, but largely as a secondary factor, i.e. it was generally a consequence of mechanical problems, disorientation, or prior mishandling of the aircraft. Personality factors also made a significant contribution, and the data suggest two distinct types of problem. Life stress and high workload appeared not to play a major part in stress-related accidents. Fatigue was not a major factor, but was closely associated with cognitive failure.

## INTRODUCTION

It is widely accepted that flying, particularly military flying, is a stressful occupation. The real significance of the stresses involved in flying is, however, not easily explicated. There are several reasons for this. First, the role of stress is equivocal. Some aviators at least are attracted by the challenge of operating under pressure of whatever kind. And the effects of stress may, under the right conditions, be beneficial. Although the inverted 'U' relationship between arousal and performance, first proposed by Yerkes and Dodson (1) eighty years ago, is by no means a full description of the complexities of stress, it is, nevertheless, a useful reminder of some salient facts: Some stressors raise arousal level, and some depress it, and either action can, at times, improve performance. In addition the experimental investigation of the effects of stress is restricted by obvious ethical and practical difficulties. As a result, the effects of relatively benign stressors in mild doses (eg fatigue, noise, hypoxia) have received attention in the laboratory and, to a lesser extent in simulations and flight tests, but one is left with the suspicion that stressors of great operational significance (particularly varieties of threat) have not yet been adequately investigated in a realistic context, despite some remarkable efforts (2).

The study of aircraft accidents offers the prospect of obtaining some clues to the operational impact of stressors and their relative importance. One may assume, perhaps with little justification but as a useful starting point, that whatever factors are found to be major causes of accidents are also likely to have a deleterious effect on operational effectiveness - perhaps in proportion to their significance in the aetiology of accidents. This gives the investigation of accidents a significance in addition to that derived from the enormous cost of individual accidents. Clues may be sought as to the origins of stress in flying, the nature of the effects of stress, and the relative importance of stress in comparison with other human factors problems.

In 1972 the Royal Air Force started a scheme allowing psychologists to conduct independent investigations of aircraft accidents in conjunction with the established Boards of Inquiry. The data discussed here were collected in the course of these investigations.

## METHODS

By the summer of 1988, 149 military flying accidents had been investigated. A few involved Royal Navy or Army aircraft; the majority were RAF accidents. The investigations drew on several sources of information:

- Confidential interviews with survivors and others.
- The personal records of those involved in the accidents.
- Eyewitness reports.
- Analysis of flight data recorder tapes, recordings of radar traces, radio transmissions etc..
- Examination of cockpit equipment, regulations, manuals and other documents.

Data on each accident were recorded in a simple computer data base. In addition to information on aircraft type, phase of flight in which the accident happened, etc., the human factors which contributed to the accident are recorded as 'possible', 'minor' or 'major' influences.

**RESULTS**

More than thirty human factors categories have been used in coding the accidents. Some form natural subgroups and have been combined into generic terms in the list in Table 1. The full list is in Appendix A. It is intuitively obvious that the factors do not all have the same logical status: Some are enabling conditions or predispositions, rather than direct causes; others describe the way in which an error occurs. An arbitrary division of the factors has been imposed on Table 1 reflecting this consideration. The three groups are: Aircrew Factors - predisposing conditions some of which are under the control of the aircrew, others being more or less natural or innate; System Factors - enabling conditions engendered by high workload, inadequacies of equipment design or training, etc.; and Modes of Failure - essentially descriptions of types of error. Table 1 shows those factors cited as at least possible contributory causes in more than 10% of the accidents. Most accident investigations revealed three or four human factors problems; some revealed ten or more.

**Table 1: The major human factors**

| | |
|---|---|
| AIRCREW FACTORS | |
| personality | 23% |
| inexperience | 20% |
| life stress | 11% |
| SYSTEM FACTORS | |
| ergonomics | 23% |
| training and briefing | 19% |
| administration | 17% |
| high workload | 14% |
| MODES OF FAILURE | |
| overarousal | 26% |
| cognitive failure | 17% |
| distraction | 16% |
| inappropriate model | 13% |
| disorientation | 13% |
| visual illusion | 12% |

A few of the terms in Table 1 require some explanation:

- Overarousal: The term 'stress' is commonly used in a variety of ways to describe both stressors and the response to them. For convenience 'overarousal' is used here to describe a non-adaptive response to stressors of an exciting or alarming nature. Similarly, 'underarousal' denotes performance degradation due to depression of arousal level.

- Life stress: Any personal or domestic events believed to have a worrying, anxiety provoking or exciting effect on an individual. The personal events may include some arising in the course of professional duties, but not, usually, short term episodes directly connected with flying.

- Administration: This term covers the content of manuals, pilot's guides, instructions and orders, and also features of chains of communication.

- Cognitive failure: A type of error in which actions fail to match intentions, usually because an intended action is omitted or because an unintended action is committed. Such failures are commonly attributed, in lay-man's terms, to 'absent-mindedness'.

- Inappropriate model: This term covers errors due to the formulation of intentions on the basis of incorrect information or assumptions.

The early accidents in the database were selected for their obvious human factors interest. The terms of reference of the scheme have changed, and now an attempt is made to investigate any accident in which aircrew error is considered to be a possible contributory cause. There are grounds, therefore, for expecting a change in the pattern of results obtained over the years. The data do not, however, fulfill this expectation. A comparison of early and late investigations reveals no significant trends.

**Origins and effects of overarousal**

Table 2 summarizes a classification of the factors chiefly responsible for a state of overarousal in the aircrew involved in the accidents, and of the effects of that overarousal on their performance. The classification was by no means easy to impose on essentially narrative data describing accidents with complex causes. It is entirely possible that some categories, such as 'disorganised response', are inflated as a result of this difficulty and that of the original investigators, who had to deal with the survivors' understandably confused recollections of alarming events. Nevertheless, the classification allows some broad distinctions to be made.

Of the 39 accidents for which overarousal was cited as a contributory factor, 19 involved a mechanical problem (such as engine failure, hydraulic or electrical failure, bird strike, lightning strike, fire or low fuel state) which was regarded as the stimulus for overarousal. In fourteen of these cases, the emergency was considered to have been in some degree mishandled, thereby increasing the danger. Precipitate and inappropriate action accounted for four cases and disorganised or slow responses for seven. Overarousal was not the only cause of mishandling of emergencies; five other cases were due to a variety of factors other than overarousal.

**Table 2: Origin and effects of acute overarousal**

| | |
|---|---|
| **Origins of overarousal:** | |
| Mechanical problems | 19[1] |
| Mishandling | 6 |
| Disorientation | 5[1] |
| Anxiety or other personality factor | 4 |
| Supervisory defects | 3 |
| Cognitive failure | 2 |
| High workload | 1 |
| **Effects of overarousal:** | |
| Disorganised response | 12 |
| Narrowing of attention | 7[2] |
| Cognitive failure | 5[2] |
| Slow response or inactivity | 4 |
| Precipitate action | 4 |
| Minor or undetermined effects | 9 |

[1] One accident included in both these categories
[2] Two accidents included in both these categories

In six accidents, overarousal followed mishandling of the aircraft. Limited talent was a predisposing factor in at least half of these.

Five accidents involved overarousal arising from disorientation. All five resulted in the loss of the aircraft. In three instances in which the pilot was killed, it is fair to say that overarousal was assumed to have been a likely concomitant of the disorientation that was believed to be the cause of the accident.

In twelve overarousal-related accidents, a crewmember's personality was thought to have been a contributory factor. Usually, (eight of the twelve) this was due to a lower than average tolerance for stress (see the section on Personality). In four accidents a predisposing personality factor was the cause of overarousal. In three of these, the origins of the overarousal lay in a crewmember's predisposition to anxiety - in one case about test sorties; in another about the possible effects of high intensity radio sources; and in a third, a general unease about fast jet flying may have been heightened and focussed on the possibility of control restrictions. The effects of overarousal in these cases were: a focussing of attention which resulted in the omission of an important action; and, in two cases, precipitate and probably unnecessary ejections.

In two accidents supervisory failings resulted directly in pilots facing novel situations with which they were ill-equipped to deal. In both cases the pilots made errors leading to their losing control of the aircraft. A third accident was similar, except that the overarousal followed the loss of control and hindered recovery; again the necessary enabling conditions included a supervisory factor.

In two accidents, problems arising from a cognitive failure caused overarousal which impeded resolution of the problems. In a further five accidents, cognitive failure appears to have been a result rather than a cause of overarousal.

**Other sources of stress**

Life stress:

In seventeen investigations it was thought relevant to record details of personal and domestic events that might have been a source stress for the aircrew involved. In eight cases overarousal was also considered to be a factor contributing to the accident. In general, however, it was not possible to make any direct link between the life stress recorded and the causes of the accident. In only two cases could personal events be viewed as having a direct causal bearing on the accident: One involved recent experience under fire, which may have caused the pilot to emphasise tactical considerations at the expense of safety; the other involved a terminated engagement to marry and subsequent rather cavalier use of an aircraft. Most of the remaining instances fall into the following groups:

- Domestic problems - five cases: deaths, illness or health problems in the family; intensive and tiring domestic activity immediately preceding the accident (two cases, also listed under fatigue).
- Marital problems - two cases: specifically worries about infidelity or incompatibility.
- Work problems - five cases (two also involve domestic stress): excessive executive responsibilities or secondary duties; conflict between domestic and professional demands.

The mode of failure for five accidents in which life stress was cited as a possible contributory factor was cognitive failure; in three cases a deliberate disregard for rules was a major factor in the accident.

Fatigue:

Although fatigue does not appear in Table 1 as a major cause of accidents, thirteen investigations (9%) did reveal fatigue as a possible contributory factor. Four accidents occurred during night flying, three of them after relatively long periods on duty. In one case night flying over the previous three nights was thought possibly to have caused fatigue on the day of the accident. In five case the fatigue originated at least partly in social or domestic activities. Cognitive failure was the main associated mode of failure (six cases); there were also two cases of apparently controlled flight into the sea, two of failure to avoid rising ground and one mid-air collision.

High workload:

Although 21 accidents implicated high workload as a contributory factor, only seven of these were associated with evidence of overarousal. Four of the seven involved mishandled emergencies, the excess workload arising from mechanical problems. Two of the remainder involved training in demanding operational conditions, which may, of themselves, have generated a degree of excitement. It is not possible to determine whether the high workload or the overarousal made the greater contribution to any of these accidents, but it may be reasonable to assume, in the four cases involving mechanical problems, that the high workload was not itself the primary cause of the overarousal.

**Other causes of accidents**

Personality:

In 34 investigations the personality of a crewmember or other relevant person was considered a possible contributory factor. Twenty cases fall into one or other of two definable sub-groups, nine in one, eleven in the other. The smaller group is characterised by comments in the subject's personal records such as: "underconfident", "nervous", "prone to over-react". Six of the nine cases involved mishandling of an emergency; one probably involved over-reaction to a mis-identified emergency. The larger group is identified by the following descriptors: "over-confident", "reckless", "disregards rules". The results of this attitude included deliberate excitement seeking (eg illegal low flying) and exhibitionism, as well as pressing on into difficulties without much thought. Two mid-air collisions and four collisions with obstructions, the ground or the sea resulted.

Supervision and ergonomics:

Poor display design accounted for 14 of the 34 accidents in which ergonomic deficiencies played a part. Nine were ascribed to poor cockpit layout and eleven to poor control design. Combining the two supervisory categories (training and briefing and administration) with the ergonomic category reveals that 65 accidents (44%) involved enabling factors generated by the system rather than by the aircrew themselves.

Cognitive failure:

Cognitive failure was a primary or contributory cause of 26 accidents. Nine of these involved actions omitted by the crew, usually from a very familiar drill; 19 involved substitution of inappropriate actions for those intended. In seven cases, distraction provoked or enabled the cognitive failure to happen. In ten cases fatigue or underarousal was considered a predisposing condition. Eight cases of cognitive failure were also associated with life stress. The most common result of cognitive failure was a wheels-up landing - ten cases in all.

**DISCUSSION**

Overarousal:

The origins of acute overarousal appear to fall into several subgroups. About half of the overarousal related accidents (13% of the total sample) involved mechanical failure, sometimes as a result of operating hazards such as birdstrikes or lightning strikes. Another important subgroup is overarousal due to disorientation. Other specific causes were problems arising from mishandling, cognitive failure or supervisory failings. Overall the first impression is of specific, single causes of overarousal, usually with a sudden onset, rather than a gradual accumulation of several minor stresses. Specific remedies might, therefore, be found in improvements in simulator training - to improve responses to emergencies - and in better presentation of attitude information. Attitude displays that address the ambient visual system rather than central vision could be of real benefit in reducing the probability of disorientation (3).

Life stress and personality:

Indications that specific, single causes of stress do not constitute the whole picture come from the data associating personality characteristics and life stress with aircrew error. Life stress has commonly been assumed to contribute to stress-related errors and has been the subject of some attention in recent years. Alkov and Borowsky (4) and Alkov et al (5) found a number of life events to be associated with involvement in aircrew error accidents. These included:

- Recent engagement to be married.
- Recent loss of a friend or relation through death.
- Marital problems.
- Recent major career decision.
- Recent trouble with peers, subordinates or senior officers.

Some additional factors seemed to be more descriptive of personality characteristics than life events:

- Lacking in maturity or stability.
- Lacking in a sense of humour concerning self.
- Experiencing difficulty with interpersonal relationships.
- Slow to assess potentially troublesome situations.
- Lacking professionalism in flying.

It is possible to interpret two of the five life events listed above (marital problems, trouble with other officers) as also reflecting immaturity or inadequacy in coping with interpersonal relations. In fact, Alkov et al interpret the findings of the two studies as indicating that social maladjustment may be a good predictor of aircrew error and they place little weight on the remaining life events. What, then, is the role of life stress? As indicated above, in only two of the 17 cases where life stress was recorded as a possibly relevant background variable was it possible to see a direct relationship between the life events and the behaviour that caused the accidents. These may be regarded as rather special cases. It is, of course, inevitable that any sizeable sample of aircrew should carry a burden of some marital disharmony, some illness, domestic upheavals and problems at work. Without a control group, it is impossible to know whether these problems are over-represented in our sample of accident victims. For the moment, the case for life stress as a direct contributor to aircrew error is, at best, not proven, and must be regarded with some suspicion until more substantial evidence becomes available. McCarron and Haakonson (6) came to a similar conclusion after surveying life events among Canadian pilots. This would probably represent the attitude of many aircrew themselves. For many the cockpit of a high performance aircraft provides a welcome refuge from down-to-earth pressures and annoyances.

The role of personality in aircrew error accidents appears to have at least two discernible aspects which account for 20 out of the 34 personality-related accidents. One aspect has a bearing on stress. Some individuals previously described by their supervisors as underconfident or nervous failed to cope when presented with emergencies or unusually demanding conditions. Precipitate, inappropriate action was a common style of error. The second group, described as overconfident or reckless, either sought excitement in unauthorised ways, or was oblivious of or slow to recognise risks. Levine et al (7) found that questionnaire items concerned with adventurousness or risk taking were associated with accident occurrences among U.S. Navy aviators. However, in a review of personality studies, Farmer (8) found that despite the existence of some evidence implicating extraversion and neuroticism, overall the evidence was inconclusive and contradictory. The two studies by Sanders and Hoffman (9) and Sanders et al (10) provide an instructive example of the difficulty of obtaining stable correlations between personality data and accident statistics. If the data presented here are any guide, it seems likely that both unstable introverts and unstable extraverts have their own idiosyncratic risks. This would certainly make it harder to demonstrate a simple correlation between extraversion/introversion, as measured by personality tests, and accident-proneness. There seems little prospect of identifying the high risk personalities with a useful degree of validity at the selection stage. However, given that supervisors are already demonstrating some awareness of relevant personality characteristics, it may be worthwhile attempting to supplement their observations with formal personality tests. These could provide the basis both of guidance for supervisors and of counselling for individuals.

Fatigue and workload:

Fatigue and high workload were both associated with relatively few stress-related accidents. It is no surprise that nearly 40% of the fatigue-related accidents involved night flying. Perhaps more interesting is the fact that domestic activities contributed to fatigue in a similar number of accidents. Both sources of fatigue should be controllable by suitable supervisory action.

Cognitive failure:

The largest homogeneous class of immediate causes of accidents appears to be cognitive failure (17%). This represents a peculiarly difficult problem to tackle, because, to a large extent, being well trained and experienced is a requirement for this type of error. Reason and Mycielska (11) found that people reporting cognitive failures were more often preoccupied (at the time of the mistake) than not, and also tended to be tired or sleepy rather than emotional or excited. There are parallels in the present data. Ten out of 26 cognitive failures were associated with fatigue or underarousal (five resulted from overarousal); eight were associated with life stress - a possible source of preoccupation. There is a more complicated link between cognitive failure and life stress, however, and one that takes account of the intuitively obvious fact that individuals differ in their response to life stress.

Broadbent et al (12) showed that proneness to cognitive failure is a relatively stable trait and that those who are prone to cognitive failure are more likely to develop minor symptoms in response to stress than those who are not. Broadbent later argued (Broadbent et al (13)) that the basis of the trait lay in differences in cognitive style, those with a more obsessional style being both less vulnerable to chronic stress and less subject to cognitive failure. He also suggested that cognitive styles become more extreme under stress. Thus, although the evidence for life stress as a direct cause of accidents is doubtful, it may have a relevance in identifying those who are most liable to cognitive failure, and, possibly, their times of highest risk. Some piecemeal remedies for cognitive failure, involving redesign of equipment, are possible. There is also a clear need for a valid, objective test of liability to cognitive failure, and for techniques of remedial training in cognitive style.

System factors:

It is a truism that complex systems, like aviation, can never be free of human error. The present data indicate that, in a substantial proportion of accidents (44%). significant errors were made by people remote from the critical events. These errors included design of equipment, inadequacies in training and briefing and administrative failures. Often the errors were not obscure or complex. Many of them were surely identifiable as potential hazards before they caused an accident. The only practical remedy for system errors of this type requires aviators to take a closer interest in the way their system operates and, perhaps more important, the relevant authorities should encourage a questioning attitude and be prepared to support changes to the system in the interests of flight safety.

## CONCLUSIONS

Although overarousal makes a significant contribution to aircrew error accidents, it appears, in general, to result less from generally high levels of stress or the cumulative effects of small stressors than from specific, provocative events. Mechanical failure and disorientation are two significant classes of provocation. Specific remedies in the form of improved simulator training and enhanced presentation of attitude information are at least conceptually feasible.

The role of life stress in accidents appears ill-defined. It seems unlikely to be a direct causal agent, and whatever significance it has may be related to some aspects of personality (social maladjustment) or cognitive style. Fatigue made a small contribution to the accidents investigated, largely in connection with night flying and, interestingly, tiring domestic activities. Nearly half the accidents involving fatigue were due to cognitive failure.

Two distinct classes of personality problem are discernible in the data. One involves overarousal in response to emergencies or other demanding circumstances, and appears to be the province of unstable introverts. The other involves excitement seeking and disregard of risks by unstable extraverts. The use of personality tests to provide guidance for supervisors and counselling for aircrew is a possible remedy.

A major cause of aircrew error was cognitive failure. Although some cognitive failures occurred in stressful conditions, they were more likely to happen in normal, undemanding circumstances, or when the aircrew were fatigued or underaroused. General remedies for this type of failure are not available and should be a priority for future research.

Nearly half of all the aircrew error accidents involved some contribution from design deficiencies, inadequacies in training or briefing, or administrative failures. Such errors represent a significant challenge for both designers of equipment and those authorities responsible for the training of aircrew and the control of flying activities.

**REFERENCES**


1. Yerkes, R. M. and Dodson, J. D. The relation of strength of stimulus to rapidity of habit formation. Journal of Comparative and Neurological Psychology, 18, 1908, 459-482.

2. Berkun, M. M. Performance decrement under psychological stress. Human Factors, 6(1), 1964, 21-30.

3. Green, R. and Farmer, E. Attitude indicators and the ambient visual system. In Report of the XVI Conference of the Western European Association for Aviation Psychology. 1985, Helsinki.

4. Alkov, R. A. and Borowsky, M. S. A questionnaire study of psychological background factors in U. S. Navy aircraft accidents. Aviation, Space and Environmental Medicine, 51(9), 1980, 860-863.

5. Alkov, R. A., Borowsky, M. S. and Gaynor, J. A. Stress coping and the U. S. Navy aircrew factor mishap. Aviation, Space and Environmental Medicine, 53(11), 1982, 1112-1115

6. McCarron, P. M. and Haakonson, N. H. Recent life change measurement in Canadian forces pilots. Aviation, Space and Environmental Medicine, 53, 1982, 6-13.

7. Levine, J. B., Lee, J. O., Ryman, D. H. and Rahe, R. H. Attitudes and accidents aboard an aircraft carrier. Aviation, Space and Environmental Medicine, 47, 1976, 82-85.

8. Farmer, E. W. Personality factors in aviation. International Journal of Aviation Safety, 2, 1984, 175-179.

9. Sanders, M. G. and Hoffman, M. A. Personality aspects of involvement in pilot-error accidents. Aviation, Space and Environmental Medicine, 46(2), 1975, 186-190.

10. Sanders, M. G., Hoffman, M. A. and Neese, T. A. Cross-validation study of the personality aspects of involvement in pilot-error accidents. Aviation, Space and Environmental Medicine, 47(2), 1976, 177-179.

11. Reason, J. and Mycielska, K. Absent-minded? The psychology of mental lapses and everyday errors. New Jersey, Prentice-Hall, 1982.

12. Broadbent, D. E., Cooper, P. F. Fitzgerald, P. and Parkes, K. R. The cognitive failures questionnaire and its correlates. British Journal of Clinical Psychology, 21, 1982, 1-16.

13. Broadbent, D. E., Broadbent, M. H. P. and Jones, J. L. Performance correlates of self-reported cognitive failure and of obsessionality. British Journal of Clinical Psychology, 25, 1986, 285-299.

## Appendix A: Human factors classification

AIRCREW FACTORS

- alcohol
- disregard for rules
- excess of zeal
- fatigue
- hypoglycaemia
- inexperience
- joie de vol (unnecessarily spirited or adventurous manoeuvring)
- lack of airmanship
- lack of talent
- life stress (exciting or worrying personal or domestic events)
- low morale
- personality
- QFI checking another QFI; reluctance to take control
- sensory limitations - visual
- social factors/crew co-ordination
- underarousal

SYSTEM FACTORS

- aircraft handling characteristics
- ergonomics - displays
- ergonomics - cockpit layout
- ergonomics - controls
- logic errors in automatic systems
- noise/communication
- operational pressures
- time pressure
- training/briefing
- administration
- physiological stress (usually heat)
- high workload
- under fire

MODES OF FAILURE

- cognitive failure - inappropriate action
- cognitive failure - omission
- disorientation
- distraction
- 'giant hand' experience
- inappropriate decision
- inappropriate model
- inappropriate spatial model
- overarousal
- slow response
- stress
- unawareness episode
- visual illusion

DISCUSSION

JONES: How common are the definitions for such terms as over arousal and life stress? What were the operational definitions of these terms and how did you measure them in such a way that you can communicate your findings with other scientists working in the same area.

CHAPPELOW: It is a problem. I have looked at other lists of factors produced by other laboratories and I don't see cognitive failure in their lists very much at all; yet I think this is our major primary cause of accidents. This may be just a question of interpretation. Life stress seems to me to be a fairly common term. What we have done was simply to record any domestic or personal events the pilot or aircrew thought fit to tell us about when we asked about it.

JONES: When you look over the series of papers that are presented here today, particularly in the morning, it comes across as if there is a reasonable amount of fragmentation and a reasonably high level of definition of the terminology that is used. Looking at the US accident data of both the Air Force and the Navy, for me the Navy, one of the common problems is the criteria for classification of accidents from the accident investigation protocols. Because these are different there is a tremendous problem when you want to talk legitimately about an approach to a solution.

CHAPPELOW: I agree with you. I think it would be very nice if we could have a common classification that would enable us to combine our accident databases. Trying to do statistics and factor analysis on 140 odd accidents is not very rewarding. It would be more worthwhile if we could combine the data from different countries. If I can offer you an excuse why we haven't done that yet, it is because I am kept very busy just investigating accidents. It would be nice to have the time to do it.

SIOMOPOULOS: It is my opinion that we do not need just one set of terms we need two. One set for the professional community of psychologists or people who investigate accidents. Another set for the people concerned with accident prevention as they are going to implement whatever we find out. The latter cannot understand our terminology, so for accident investigation we need an agreed scientific terminology and we need another set of agreed common terms for the people who implement the changes.

# REVUE DES ACCIDENTS AERIENS GRAVES
# A LA FORCE AERIENNE BELGE
# DEPUIS SEPTEMBRE 1973 - JUIN 1986

## Causes et comparaison avec quelques données de sélection

J-C. GENON *, P. TECK ** and P. VANDENBOSCH ***

* Cpt J-C. GENON, Chief- psychologist at the Belgian Medical Aerospace Center.

** Cpt P. TECK, Instructor-Pilot at the 5th squadron and selection officer at the Belgian Medical Aerospace Center.

*** Col P. VANDENBOSCH, M.D., Director Aeromedical Services, BAF.

King Albert 1 Quarters, Air Force Base, Rue de la Fusée 70, B-1130 - BRUSSELS, BELGIUM.

1. Méthode.

La Force Aérienne Belge maintient un organisme permanent, le "SEAA" (1), dont la mission consiste à investiguer lors de tout accident grave afin de proposer à l'autorité des mesures de prévention.

Cette équipe de spécialistes permanents enquête dans les domaines technologiques, météorologiques aéronautiques et médicaux.

Chaque enquête se conclut par la rédaction d'un rapport détaillé dans lequel les causes avérées probables ou possibles sont incluses.

Il est à noter que, faute d'un personnel d'investigation suffisant, aucune intervention systématique n'a été prévue dans le domaine psychologique.
Un premier pas a été franchi depuis deux ans par la décision d'un entretien systématique après un accident aérien mais seulement dans le but de procurer une expérience clinique aux psychologues appelés le cas échéant à fournir un avis. De plus, c'est le droit du président du SEAA de commettre un psychologue lors de cas particulier.

Compte tenu de la situation présente, la méthode employée pour procéder à une revue des accidents survenus sur une période de treize ans se déroule comme suit :

1° Analyse globale des dossiers d'accidents par une équipe restreinte :
1 psychologue d'aviation et 1 pilote expérimenté.

2° De cette première approche, détermination d'une base d'analyse systématique des dossiers.

3° Analyse systématique des mêmes dossiers en fonction de la base choisie auparavant.

Notre approche s'est d'autre part fondée sur le principe suivant :
compte tenu de la difficulté, pour une commission d'enquête, de connaître les facteurs déterminant les accidents, il a été retenu les éléments susceptibles d'avoir joué un rôle avec une probabilité raisonnable.

Il n'y a, dans notre recherche, aucune prétention à l'emploi d'une quelconque technique "exacte" ou "scientifique". D'emblée, nous nous sommes placés dans une optique empirique et clinique ce qui ne nous apparaît pas comme un défaut. D'autant plus que, dans le domaine des causes d'accident , il y a lieu de montrer la plus grande prudence devant la complexité des processus mis en jeu.

Une dernière considération préalable concerne l'expression "facteur humain". Même si notre acceptation de cette expression n'est pas nouvelle, il est utile de rappeler qu'elle n'implique aucune culpabilisation des pilotes concernés. Il est humain que l'évocation d'une erreur commise entraîne une réaction narcissique de la part du corps des pilotes.

2. Données de base.

Les dossiers d'accidents pris en considération ont fourni une base d'investigation décrite dans les tableaux 1 à 4 repris ci-dessous.

Il convient de noter que la quasi totalité des enquêtes a donné lieu à l'établissement de causes ou du moins d'hypothèses probables.
D'autre part, presqu'un tiers des accidents implique deux pilotes soit parce qu'il s'agissait d'une collision en vol soit parce que l'avion était un biplace. Compte tenu du nombre limité d'avions de ce type, nous soulignons la fréquence de cet élément sans pouvoir l'expliquer plus avant.

Tableau 1.

| Investigation | % |
|---|---|
| Causes connues, possibles, probables | 99 |
| Causes inconnues totalement | 1 |

Tableau 2

| Circonstances | % |
|---|---|
| Chute | 58 |
| Collision (en vol | 10 |
| Atterrissage | 15 |
| Décollage | 11 |
| Collission (objet) | 3 |
| Collission (obstacle) | 3 |

Tableau 3.

| Types d'avion | % |
|---|---|
| Mirage V B | 42 |
| F 104 G / TF-104-G | 21 |
| F 16 A/B | 13 |
| T 33, A Jet, CM 170 | 14 |
| SV 4 bis, SF-260-M | 9 |
| Merlin | 1 |

Tableau 4.

| Pilotes impliqués par accident | % |
|---|---|
| 1 pilote | 73 |
| 2 pilotes | 27 |
| - Collisions | 10 |
| - biplace | 17 |

3. <u>Facteurs générateurs d'accidents</u>

3.1. <u>Conjugaison de facteurs</u>

Comme il a déjà été souvent souligné, la plupart des accidents découlent de plusieurs sources. Il est probable que le nombre d'accidents recensés ici et qui ne sont rapportés qu'à un seul facteur est surestimé suite aux limites des enquêtes.

Tableau 5.

| Combinaison de facteurs | % |
|---|---|
| 1 facteur | 41 |
| 2 facteurs | 44 |
| 3 facteurs | 15 |
| TOTAL | 100 |

3.2. <u>Nature des facteurs</u>

Trois grandes catégories de facteurs d'accidents ont été isolées :

- le facteur humain, à savoir tout ce qui se rapporte au pilote.
- l'inexpérience.
- les facteurs extérieurs c'est-à-dire tout ce qui ne concerne pas le pilote.

l'importance respective de ces trois facteurs, lorsqu'ils sont rapportés isolément est la suivante :

Tableau 6 :

| Facteurs isolés | % |
|---|---|
| Facteur humain | 16 |
| Inexpérience | 1 |
| Facteur extérieur | 24 |
| Total | 41/100 |

Lorsque 2 ou 3 facteurs se conjuguent, nous obtenons la répartition suivante :

Tableau 7 :

| Deux facteurs conjugués | % |
|---|---|
| Facteur humain + Inexpérience | 22 |
| Inexpérience + Facteur extérieur | 1 |
| Facteur humain + facteur extérieur | 21 |
| Total | 44/100 |

Tableau 8 :

| Trois facteurs conjugués | % |
|---|---|
| Facteur humain + Inexpérience + Facteur extérieur | 15 |

Sur l'ensemble des accidents, l'existence d'une influence des trois facteurs prend les les proportions suivantes :

Tableau 9 :

| Ensemble des accidents (facteurs isolés et conjugués par 2 ou 3) | % * |
|---|---|
| Facteur humain | 73 |
| Inexpérience | 39 |
| Facteur extérieur | 61 |

(* en fonction des 91 accidents)

Les tableaux 6 à 9 permettent les commentaires suivants :

- les facteurs humains sont ceux qui reviennent avec la plus grande fréquence mais lorsqu'ils opèrent seuls, leur importance reste limitée.
- les facteurs extérieurs ne doivent pas être sous estimés. Au contraire, ils constituent souvent le point de départ du processus qui mènera à l'accident.
- l'inexpérience est une source d'accident plus fréquent qu'on pourrait le croire.
- enfin, mais c'est là une évidence, l'inexpérience entraîne la mise en oeuvre de l'un ou l'autre facteur humain. D'où l'existence d'une forte liaison entre ces deux éléments.

## 3.3. Les facteurs humains.

Tableau 10

| Facteurs humains * | % ** |
|---|---|
| Domaine perceptif | 12,1 |
| Pèrception visuelle | 05,5 |
| Désorientation | 06,6 |
| Domaine médical | 7,7 |
| Visus | 01,1 |
| Psychique | 03,3 |
| Autre | 03,3 |
| Domaine psychique | 33 |
| Surestimation (soi) | 11,0 |
| Emotionalité | 17,6 |
| Fatigue | 04,4 |
| Domaine comportemental | 15,4 |
| Faute | 12,1 |
| Toxicomanie | 03,3 |
| Domaine Aéronautique | 70,4 |
| Maîtrise des paramètres | 22,0 |
| Vigilance/Attention dispersée | 14,3 |
| Exécution des procédures | 30,8 |
| Communications | 03,3 |

(* Certains facteurs peuvent exister en conjugaison avec d'autres).

(** en fonction des 91 accidents).

L'analyse des différents facteurs humains en jeu montre que :

- les éléments perceptifs sont relativement faibles. Il est cependant à noter que la "désorientation" est ici prise dans un sens très restrictif qui ne se rapporte qu'à la localisation, par le pilote, de lui même et de son avion dans l'espace à trois dimensions, à l'exclusion des vertiges ou des erreurs de pilotage.
- les éléments médicaux sont les moins importants ce qui témoignent de l'efficacité des sélections et des suivis médicaux ainsi que de la nécessité à maintenir des seuils d'exigence élevés.
- dans un tiers des accidents, l'affectivité joue un rôle primordial.
  Souligner la présence du stress, en particulier lorsque des problèmes (techniques, météo...) se posent, n'est pas d'une très grande originalité. Mais, justement, c'est à cause de ce facteur inhérent au vol qu'il y a lieu de prévoir certaines mesures, en particulier dans les exigences de l'entraînement (en vol <u>et</u> au simulateur) pour le compenser.
  Quant à la "surestimation de soi", on ne répétera jamais assez que le meilleur pilote du monde n'est pas le meilleur tous les jours et qu'un minimum d'insight, pour un pilote, est une excellente prophylaxie.
  Autrement dit, si aucun pilote ne "sortait" de son enveloppe personnelle, on éviterait un accident sur dix.
- il en va de même pour les imprudences (les "fautes" dans le tableau 10).
  Le taux est bas car la toute grande majorité des pilotes sont de vrais professionnels mais, justement, quinze pourcents d'accidents, souvent mortels, pour être sorti de l'enveloppe de vol ou technique de l'avion c'est encore trop.
- dans le domaine, dit ici "aéronautique", on retrouve quatre des grandes activités du vol. Comme celle ci sont la traduction dernière de tout le fonctionnement psychique il est normal que l'importance de cet élément soit élevée.
- la désorientation nous a amené à faire une comparaison entre deux générations d'avions. Les valeurs absolues sont trop faibles pour tirer des conclusions mais il y a peut être là un problème pour l'avenir.

Tableau 11

| Désorientation | % |
|---|---|
| F 104 G/MIRAGE V B | 08,6 |
| F 16 A/B | 16,7 |

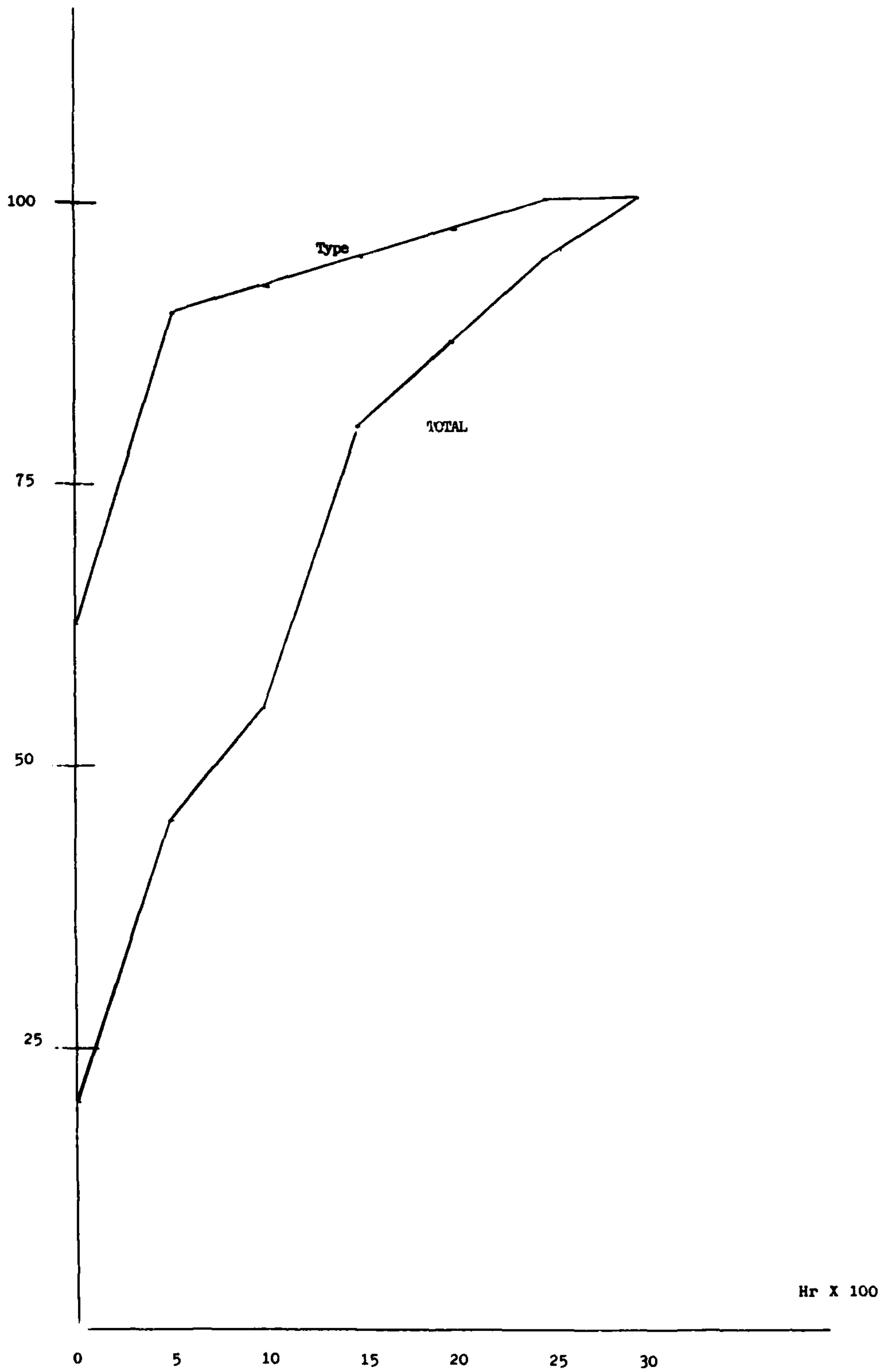
3.4. <u>L'inexpérience</u>

Tableau 12

| Inexpérience | % |
|---|---|
| Apprentissage | 05,5 |
| Conversion | 01,1 |
| Début | 25,3 |
| Interruptions | 07,7 |
| TOTAL | 39,6/100 |

Il est dit que pour bien faire quelque chose, il faut pratiquer souvent. L'élève est bien encadré, il court peu de risque. Mais presqu'un accident sur dix arrive à des pilotes que leur fonction écarte du vol. D'autre part, si un quart des accidents arrive dans les débuts, il faut s'entendre sur ce mot. En effet, le graphique 1 montre que l'expérience totale est moins concernée que l'expérience sur le type d'avion. Il ne faut pas exagérer l'interprétation des deux courbes car plusieurs éléments indéterminés doivent jouer (par exemple, le nombre de pilotes expérimentés) mais le transfert d'un avion à un autre est un moment sensible et peut-être le retour à partir d'un avion sophistiqué vers un avion plus classique l'est-il encore plus.

Graphique 1 :



### 3.5. Les facteurs extérieurs.

Tableau 13

| Facteurs extérieurs * | % |
|---|---|
| Incidents techniques | 38,5 |
| Conditions de vol | 24,2 |
| Chocs | 11,0 |
| TOTAL | 73,7 /100 |

(* certains facteurs peuvent exister en conjugaison avec d'autres)

Il est de mode de souligner le rôle du facteur humain. Même s'il est important, il ne doit cependant pas occulter l'importance des conditions matérielles. Les problèmes techniques, même si parfois, ils ont une origine humaine dans la maintenance, existent ! il en va de même pour les conditions météorologiques et les exigences des missions (par exemple, vol à basse altitude, exercices, compétitions internationales).

## 4. Liaison avec quelques données de sélection

Une sélection est un processus qui évolue avec le temps. Pour couvrir l'entièreté de notre population nous n'avons retenu, pour chaque aspect, que trois grandes catégories :

- Faible - Défavorable (-)
- Moyen (±)
- Bon - Favorable (+)

Le but n'est pas de décrire ni de défendre un système de sélection. Mais de dégager certaines tendances au sein de notre population que nous formulerons sous forme de questions.

### 4.1. Pronostics.

Tableau 14

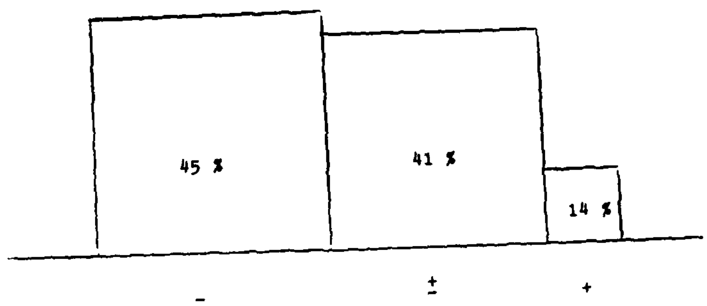
| | | Personnalité - | Personnalité ± | Personnalité + |
|---|---|---|---|---|
| Psychomotricité | - | * 31,4 | 17,1 | 08,6 |
| | ± | 03,8 | 08,6 | 03,8 |
| | + | 08,6 | 10,5 | 07,6 |

(* % d'individus)

Un mauvais score en psychomotricité couplé avec un jugement défavorable de la personnalité prédisposent-ils à l'accident ?

4.2. Psychomotricité

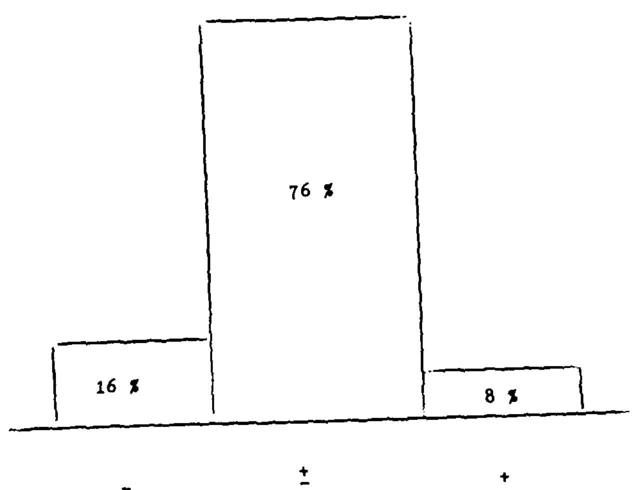
Graphique 2 :




Du graphique 2, un aspect semble se dégager : le nombre de sujets présentant un score faible pour l'ensemble de la psychomotricité, tend à être notable.

4.3. Représentation spatiale.

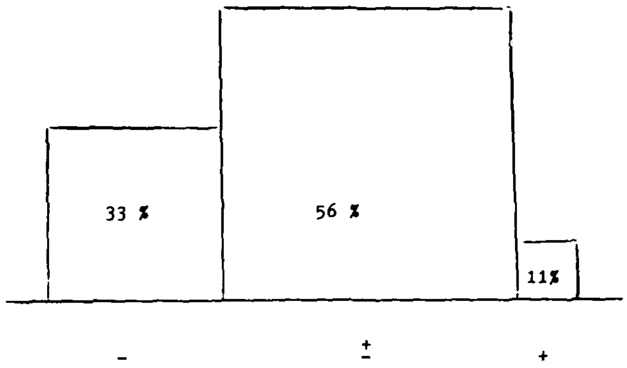
Graphique 3 :




Aucune tendance nette ne se dégage.

4.4. Coordination

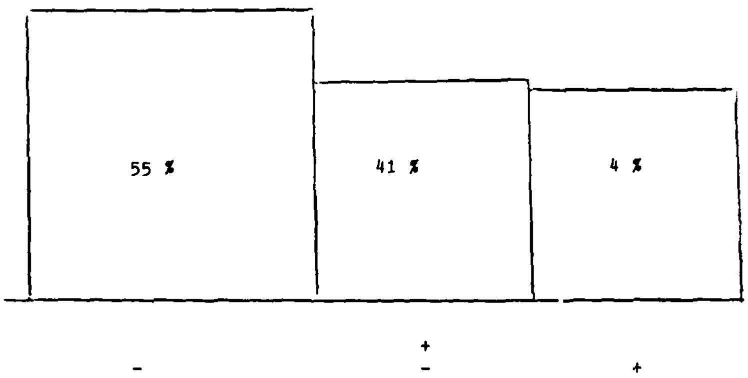
Graphique 4 :




Une tâche simple de coordination vue-mains/pieds permet-elle une certaine prédiction à long terme quant à l'efficacité en vol ?

4.5. Réaction de choix (Temps de réaction).

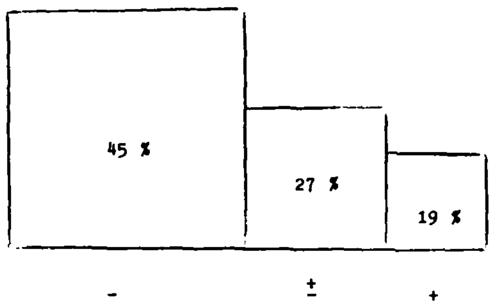
Graphique 5 :




En situation de laboratoire, lors d'une tâche de réaction de choix, le délai nécessaire à la réaction motrice est-il lié à des caractéristiques neuro-physiologiques susceptibles d'être un facteur de risque?.

### 4.6. Personnalité

Graphique 6 :




Inconnus : 10

Une sélection négative fondée sur l'investigation de la personnalité peut-elle conduire, si le seuil d'exigence est assez élevé, à une réduction des accidents?

### 4.7. Données complémentaires.

Tableau 15 :

| EEG et / ou Antécédents |
| --- |
| % |
| 10,9 |

Un sujet sur dix présentait un doute soit au niveau de l'électroencéphalogramme soit dans l'anamnèse.

Tableau 16 :

| Accidents connus |
| --- |
| % |
| 20,9 |

Avant l'accident aérien, un cinquième des sujets avait eu un incident ou un accident dans les domaines de l'aviation ou de l'automobile.

## 5. Conclusions

Une enquête sur un accident est a fortiori une tâche complexe qui, très souvent, ne peut être meneé complètement à bien et qui entraîne autant d'hypothèses que de certitudes. A fortiori, une recherche sur dossier doit, de par la méthode, s'imposer des buts limités.

D'autre part à la limite, chaque accident est unique. Le sujet se prête à la systémisation. Pourtant la prévention demande le dégagement d'éléments communs.

Et, en fait, ces éléments communs reviennent sans cesse. Même notre approche empirique les met en évidence, tout au moins jusqu'à un certain niveau d'investigation. Malheureusement, ce sont souvent les choses les plus évidentes qu'il faut répéter sans cesse pour obtenir une bonne prévention.

En synthèse, dans le domaine médico-psychologique, nous retiendrons les exigences suivantes:

- un haut niveau d'exigence lors des sélections médicale et psychologique.
- un haut niveau d'exigence lors des entraînements et des conversions.
- un haut niveau d'exigence dans la maintenance et le suivi technique des appareils.
- un suivi médico-psychologique sérieux qui est d'abord une responsabilité des pilotes eux-mêmes et ensuite l'oeuvre des différents spécialistes.

# Accidents in fighter aircraft caused by "human factors". Why do they occur?

by

G. Myhre

RNoAF Institute of Aviation Medicine
P.O. Box 14, Blindern
OSLO, NORWAY

Pilots in fighter aircraft represent a rigorously selected group both physically and psychologically. Still, more than half of the incidents and accidents involving these aircraft can be attributed to human factor overloading, even during routine operations. In what way is this high number of human error accidents explained considering the fact that the psychological selection tests have never been more sophisticated and thorough than they are today?

The Typical fighter pilot is characterized by that

- he is capable, competitive, individualistic, scores high on achievement and exhibition.
- his values are capitalistic rather than radical
- he is intelligent, athletic and blessed with good health
- he lives an active life, has many acquintances but few close friends
- he is well adapted and independent, prefers physical activity to intellectual ones
- he has limited contact with his own feelings and is not prone to introspection.

In summary: The typical abilities of the fighter pilot have been exactly what Hollywood has stressed in the movies about pilots: The masculine man, or what has been defined as masculinity through western history.
have selection procedures for modern fighter pilots relevance to what is required of this species of mankind.

Selection is procedures based on the recognition of the fact that some individuals are better suited for some tasks than others, and that there are more people who want to do these tasks than required. During WW I aviation was in its beginning, and men who could handle an air plane were picked. It was not until WW II efforts were made to develope tests for candidates wanting to become aviators. Such selection should assure minimum losses during training and combat, and maximize the effectiveness in an operational setting. This selection procedure has been further improved after WW II. Concomitantly a reduction of aviation accidents were registred. This is probably due to a combination of selection procedures, improved training and improved technical standards of the aircraft.

In Norway the applicants for undergraduate pilot training are psychologically tested for:

- mechanical skills
- spatial perception
- perceptual speed
- time evaluation ability
- arithmetics, mathematical understanding
- performance with instrumental panels
- personality (interviews and tests with emphasize on psychological defence mechanisms).

This testbattery is supposed to test the qualifications considered important in relation to the requirements of the pilot in a fighter cockpit. Nevertheless, even pilots become vulnerable when they are under pressure. What are the causes, and how do we mitigate them?

One possibility to study these causes is to participate in accident investigations and afterward analyzing the accident factors, mitigate the causes by preventive work.

When an aviation occurence was due to human error earlier there was a tendency to go no further than stating this judgement. It was common belief that human fallability was not modifiable. When a mistake was made, however, it was recognized as an isolated occurrence with no connection to other incidents or accidents.

Human factor as a cause in aviation mishap may be subdivided into the following diciplines:

A PHYSIOLOGY. The effects of flying on the human body

B PSYCHOLOGY. Emotions, motivation and information processing

C SOCIOLOGY. Social relationships, personality and life event stresses

D PATHOLOGY. The condition of the body before, during and after the impact

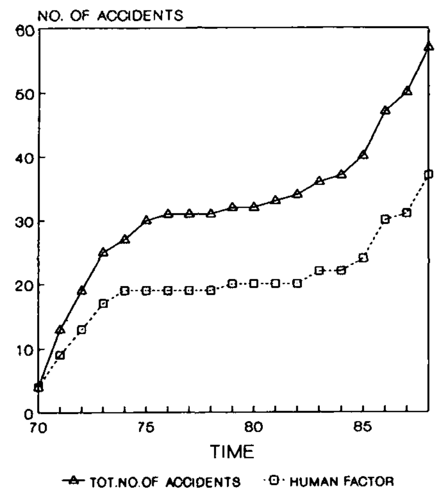
Each of these diciplines must be examined in each case: How was the pilots vigilance at the time of the mishap/accident? Was he tired, exhausted? Questions like these are relevant because physiological factors and limitations will interact with and affect the pilot's cognitive processes, perception and memory. As a requisite preparation for accident prevention it is therefore important to inform the pilot of his vulnerability within the above mentioned disciplines, even if they belong to a selected group considering certain psychological and physiological qualifications.

In Norway this fact has lead to courses in aviation medicine and psychology where the aim is to inform about a healthy way of living including suggestions for physical excercise and nutrition. How lack of sleep and irregular meals may affect the performance. Further the pilots are instructed how psychological factors and change in cognitive capacity may reduce the awareness, and that social relationship or lack of it will be reflected in their work.These aspects have been accepted as important for the professional development of the aviator.

In spite of this new information that is beeing transmitted to the aviators why do not the number of accidents caused by human factor whittle down?

Looking at accidents for F-5 and F-16 in Norway from 1970 up to 1988 we have had 57 board investigated accidents where 33 were categorized as being caused by human factor e.i. close to 60% of these cases 37 involved the F-5, and 13 the F-16. Of these 24 or 65% of the F-5 accidents were classified as human error, and 11 or 61% of the F-16 accidents. Figure I illustrates how this trend is distributed over these 18 years. It is striking to see how the human factor overloading accidentsare increasing and that this trend has continued through 1988.

FIGHTER AIRCRAFT ACCIDENTS




The term "pilot error" as a cause of aviation accidents should be avoided. This term is incorrect, it only describes the <u>outcome</u> of several unfortunate factors working together. The cause of an accident is to be found in what interfered with the pilot's judgement at a critical moment.
Table I shows three categories of causes of accidents.

FIGHTER AIRCRAFT ACCIDENTS

| YEAR | CAUSE OF ACCIDENT | | | |
|---|---|---|---|---|
| | TECHNICAL FACTORS | INADEQUATE TRAINING/BRIEFING | STRESS PRIOR TO FLIGHT | UNKNOWN |
| 1970 | 2 | | 2 | |
| 1971 | 4 | 2 | 3 | |
| 1972 | 2 | 2 | 2 | |
| 1973 | 2 | 1 | 3 | |
| 1974 | | 2 | | |
| 1975 | 1 | | | 2 |
| 1976 | | | | 1 |
| 1977 | | | | |
| 1978 | | | | |
| 1979 | | 1 | | |
| 1980 | | | | |
| 1981 | 1 | | | |
| 1982 | 1 | | | |
| 1983 | | 1 | 1 | |
| 1984 | 1 | | | |
| 1985 | | | 2 | 1 |
| 1986 | 2 | 4 | 1 | |
| 1987 | 2 | | 1 | |
| 1988 | 1 | 3 | 2 | 1 |

It is apparent that the number of accidents caused by technical factors remain fairly stable over the last 15 years except for a period between 1975 and 1981 when there were no accidents/mishaps large enough to require an accident board! Accidents caused by human factors display a peak early in the 1970 and demonstrate the same trend from 1985.

Looking into the causes that are hidden behind the heading of: "Inadequate training, briefing" we find pilots with little experience on the actual aircraft, and with an inclination to return to old well established procedures which were correct in their former aircraft, but which prove fatal on the new one. This trend is named both proactive and retroactive inhibition.

Such a phenomenon should be stressed by flight instructors to their students. Transfer of training is a central aspect in all learning situations. There are great individual differences in the ability of transferring from one situation to another. The flight instructors should stress this fact about interference of proactive and retroactive inhibition with new learning to transfer pilots. They should also in their briefings stress that no flight is "just routine". The attitude that every flight contains one or several new elements is an important way of preventing the pilot to believe there are sequences during a flight when he may allow himself to be less alert. The instructor should also make sure that the student pilots understand that correct, well established procedures which become integrated will function also when a deviation is in progress, and thus leave the pilot with more spare capacity in an emergency situation.

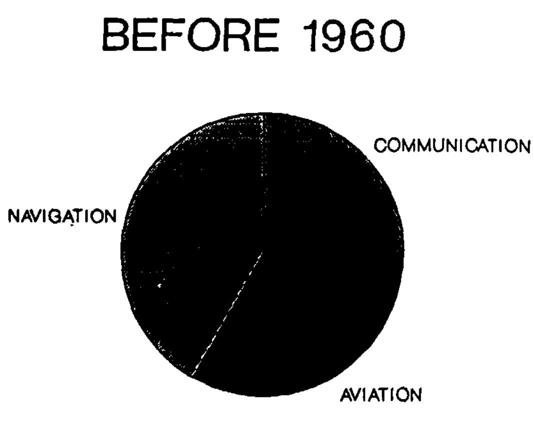
"Stress prior to flight" is also a heading under human factor accidents causes. Alongside the technical briefing of a flight a psychological briefing should follow. The pilot should be made aware how his own psychological state of mind is. In 17 of the accidents in our material the cause of the accident/mishap were due to pilot stress prior to the flight. Such stress may be concealed in changes in the family situation, pleasant ones (like getting a new baby) or unpleasant ones (the pilot is getting divorced). Both life changes are representing stress that may interfere with the pilot's flight performance.

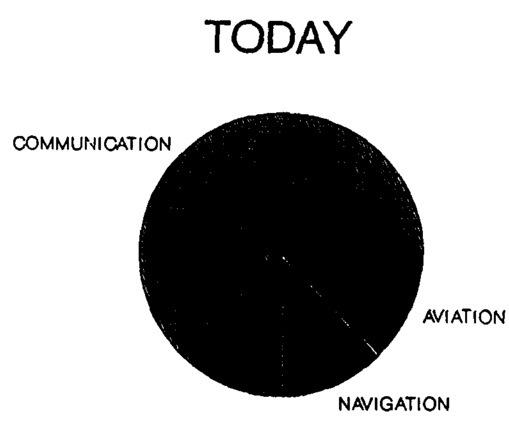
Among the accidents investigated in this study, five of the pilots reported that they had passed a test or an exam the day prior to the accident. A passed test may induce a feeling of relaxation in the pilot which is incompatible with agility and alertness. Everything turn out all right as long as the flight proceeds according to the briefing. A sudden distraction may prove hazardious, because the "relaxed" pilot's attention deficit is revealed and an accident is inevitable.

During the five last years I have talked to the surviving pilots involved in the above reported accidents. These talks revealed in most of the pilots a feeling of insecurity about their ability to perform according to instructions in the future flights and an insecurity with their whole situation. Suggestions were therefore made that after an accident a pilot should proceed his training at a lower level than prior to the accident in order to acquire the same level of confidence he had prior to the event/ mishap.

What remain still a puzzle is how this increase in human factor accidents may be explained in spite of the psychological tests, have never been more sophisticated and thorough?

Lt Col Heijboer of the RNLAF has illustrated the change in demands on a military aviator from WW I and today in stating that the pilot's main task during WW I was to aviate, then navigate and last communicate. The pilot's workload distribution today is mainly communication, then navigation and aviation. This change may mainly be explained by the development of the modern aircrafts.



Could this change in priorities explain why even after thorough psychological and physiological testing that the main fraction of accidents are still caused by human error? Do we need some additional tests in order to meet new requirements?

Let us take a look at aviation history. During the world wars it was known that in order to perform and still survive many of the pilots made procedural shortcuts. This was tacitly accepted in the squadron both by the authorities and by the pilot's peers. The planes were not as sophisticated as today, there were room for individual improvisation, and still possible to end up safely and according to the preflight briefing.

The procedures were accepted especially since they were initiated by the authorities, people with higher ranks. The ability to improvise is also one ability that was accepted as positive for a pilot, since that could be the only mean to survive in certain critical situations.

Has something changed lately that may explain this discrepancy between test results and practice? Could this be a question about masculine roles versus safety?

In the western world men must prove themselves through competition like gallinaceous birds, and some animals: The toughest male gets the best reputation in the men's world, just as the strongest bird gets the best territory and the best hen in the bird's world and the toughest lion gets the finest lioness. This "attractive" toughness is not learned in any school. It is probably accuired among the boys since I am told that you don't pick on the boy who is supposed to be strong and tough, unless you want trouble. It can also be learned just by reputation and from movies of heroes, and from older peers with lots of ideas, but with little solid information. Masculine behavior, unlike in animals, is not inborn to man, only the genes which determine the sex. Society determines in what way the different sexes should behave, and to some masculinity may be domonstrated by the ability to land and aircraft under minimum conditions or by wise decision to choose another airfield.

If we stick to the comparison of territorial animals and the military system we find that both belong to fairly strict constructed hierarchies. A hierarchy according to Webster's new world dictionary is a group of individuals arranged in order of rank, grade etc. We know now that a hierarchy makes it easy for the group members to recognise modes of behavior and identify each other by just looking at certain features. In birds these features are feather condition, color and combsize, while within the military system the group members are identified by bars, stripes and stars. When the rank of a person is established one knows how to behave towards him, as in the bird hierarchies. A hierarchy is maintained by a set of rules which must be accepted if the system shall function adequately. A hierarchy is dependent upon the group members' attitude and knowledge of the system. When everybody within the system know their rank and act according to the rules, everything functions in compliance with its purpose; that is to withstand enemies and not spend unneccessary energy identifying the members of different hierarchies.

Without embarrassing anybody, this is also the way the system is supposed to work at squadron level: the ranking order is known, the rules are accepted and the pilot's whole attention is on the flying missions and their purposes. Or is it?

During the last 20 years there has been a steadily decrease in general respect for rules and regulations in society. Within military ranks the decrement has been slower than in society as a whole thanks to a rigid system. This disrespect may have infiltrated the pilots' attitude towards preflight briefings and standing procedures and led him to make shortcuts, using his improvising ability. Fortunately most of the time the pilots, because of their skill and aptitude for improvisation, have landed safely on the airfield and the mission has been recorded as "uneventful". Such improvisation has also been possible to a certain extent in the F-5, since the demand for communication is not as heavy as in the F-16. Because of the greater complexity of the cockpit and of the pilots' task in this plane, there is no room for shortcuts of procedures which can be saved by inventive brains. This special work situation combined with a general trend to overlook and not respect rules and regulations, may actually represent the problems we are here to discuss.

All military Forces have pilots with attitudes like: My private procedures are better than the standing ones. Or: The standing procedures are for the average pilot, while I as a flying ace am allowed to make short cuts. Or: I want to object to authority. Or: I am sure nobody will notice this little shortcut which I have to make because I had my mind on something else for a moment. This last attitude may as well be unconcious as concious.

In Norway today the pilot students are not tested on attitude towards authorities, rules and regulations. Their general mentality is tried mapped during an interview. Most of the times the applicants may conceal their standing on matters like authorities and rules during an interview, while I feel that an attitude test may reveal unwanted personality traits more effectively and thereby ensuring that constant training without improvisation on effective procedures, will make them become an integral way of life. The aviator's additional training on emergency procedures will imply that he can assess when a deviation is in progress as well as know the remedial actions to take, and really do it because he has been selected also for his willingness to accept rules and regulations.

People will continue to make human errors that will lead to accidents but I think that to inform the pilots and make them aware of their weak points for instance together with the preflight briefing, and at the same time be sure the selection procedures also take care of the applicants' general attitudes towards accepting rules and procedures, we may be able to reduce much of the potential for future accidents.

## DISCUSSION

SIOMOPOULOS: Initially, I would like to raise an objection about the matter of respecting rules and authority which stems from my experience in my own Air Force. We have to keep in mind that I am talking about a particular culture, but in my own country what I have noticed is exactly the opposite. Pilots of the first jet era would do the sort of things you are talking about. Younger pilots are not like that. They are not as masculine as you describe. They do not want to stick their neck out. They are not doing the kind of things that people in the past would do. The modern aircraft does not really permit these kind of masculine "games".

My second point is that you cannot test the attitude of a youngster who is 17, 18 yr old against authority and decide on the basis of what you find if he is going to be a good pilot or not. If you have a youngster of 17 or 18 yr who respects authority then you may be dealing with an adolescent who has foreclosure of his identity. He may be an inadequate personality who uses rules and regulations and complies with authority in order to cover his inadequacy. You do not need that kind of pilot. Again the issue of deciding how we go about these things transculturally for universals would be useful.

I also have something to say about a female talking to males. Were men really taught to be male and how did they perceive the masculine identity? Now I think that only males can talk about that; females can talk about whatever has been taught to them for their female role. I think that talking about competition and raising these kind of issues is contentious. I do not know, maybe Diana would fight.

The area that we are trying to investigate is very slippery, so we have to try and find some hard data. A flight (sortie) does not happen in a vacuum. It has geographical dimensions plus temporal dimensions so when it is investigated a decision must be made on what to focus on; and of course cultural factors always have to be kept in mind.

MYHRE: May I answer some of these questions? You accused me of being a female looking into a man's world but attitudes in Greece and Norway are probably very different in this respect. I talk to pilots a lot and I feel that they can tell me more than they may dare to tell you as a man. There are still things that one should look into. In particular one should not neglect to talk to pilots about this problem of attitudes and procedures. They admit that there is a problem today because they are probably raised in a very different way from pilots in Greece. Our pilots do not care when it comes to procedures, and do not respect their parents in the way they do in Greece. So I think there are quite a lot of differences which should be looked into, even if they are not the same all over the world; but I am pleased that there is some discussion about this.

STRONGIN: I have a brief comment to add to the previous one and a question. The comment is that we have several female aviation psychologists who have achieved considerable success in working with squadron personnel. The question I have is whether you have considered, as we have to some extent, training the lower level operational leaders, the flight leaders, in the use of these kinds of principles? Does knowing about personality, knowing about the manifestation of personality in aircrew, help that level of the leadership take advantage of the unique personality characteristics of flyers. In particular we want respect for authority that can be achieved either by selecting people who willingly submit to authority, or as you know, improving the quality of leadership so that the same person will submit himself to the authority of a superb leader.

MYHRE: We are trying to do that and fortunately quite a lot of the lower level leaders in squadrons in Norway are very interested in learning about these things; we have a very good communication going on.

STOKES: If you associate increased accident rates with non-conformity we might expect to see some change in the statistics, not so much in the conservative 70's and 80's, but perhaps in the 1960's. Did we actually find that?

MYHRE: Unfortunately, I didn't look into it, but the hippies were not as active in Norway as in the States at that time, so I don't know.

STOKES: As a follow up to that, I understand that the Warsaw Pact forces are often associated with certain doctrinal rigidity and a lot of respect for hierarchy. Is it something we would want to emulate?

MYHRE: I think something in between what it is today and what it has been.

THACKRAY: Looking at the figure you presented, the dramatic rise in accidents appears to have occurred from 1985 up to the present. One would speculate that it would continue, yet I am not aware that there has been so much of a major shift in attitudes towards lack of compliance and respect for authority during this particular time period. Is this my misinterpreting your figure? Would you suggest another cause for the dramatic rise in the accident rate?

MYHRE: I might explain this by saying that the accident rate rose after we started with the F-16s. People could improvise a little more with the F-5s. That is my explanation.

STEVENSON: I do not have the figure in my head, but if I went back and looked I would have to question whether or not part of the increase in human factors mishaps is not due to 1) the decrease in mechanical problems with aeroplanes as we become more familar with the new models where we have corrected some of the faults; 2) more diverse missions. Instead of expecting an aircraft to do one thing we now expect it to be a high level interceptor, a low level attack bomber and in some cases a nuclear strike aircraft. These roles have put the aviator much more at risk than fighter pilots in the past. It seems to me that it might explain much of the increase in the human factors as a cause. When we look at the pie chart we notice one area increasing, but what happened to the other areas of the pie chart to make them go down.

MYHRE: Yes, that is right, but still the causes of accidents, labelled human factors, are the same during this period.

# PSYCHOLOGICAL REACTIONS OF PILOTS INVOLVED IN ACCIDENTS IN THE SWEDISH AIR FORCE

by

Kristina Pollack, Chief Psychologist
Flygsakerhetsinspektionen
Flygstaben
Box 80004
104 50 Stockholm, Sweden

## ABSTRACT

Every aircraft accident could be described as a unique life-event, complex in nature with varying causes and effects. Studies of the emotional consequences of accidents in military aircrew are sparse. The aim of this study was to obtain data from pilots involved in accidents in the Swedish Armed Forces concerning their psychological reactions to accidents.

As a part of the air safety program in the Swedish Air Force retrospective information was obtained from 40 pilots who had survived military aircraft crashes during 1978-85. Six were pilots in the Swedish Army and the remaining 34 were SAF pilots, 29 of whom had ejected from high-performance aircrafts. The present study covers the 34 SAF pilots only.

In order to collect objective and subjective data all pilots completed comprehensive questionnaire covering the following areas:

1) the air crash

2) the course of events during the ejection and the following rescue

3) the medical consequences and the emotional sequels, i e reactions, thoughts and mood after survival

4) attitudes towards resuming flying duty

Free comments were encouraged in the responses to questions concerning desired psychological support and rehabilitation. All pilots were asked to rate their emotional experience in the Life Change-Scale devised by Holmes and Rahe.

Two thirds of the 34 SAF pilots had returned to flying duty within 30 days of their accidents - one third was already flying again within a week. The remaining third, which took longer than one month to return to flying, felt mentally prepared to fly but were prevented from doing so - mainly by medical injuries. Three pilots (9 per cent) did not resume flying. Although some pilots rated the accident as a significantly stressful event on the Life-Change Scale, the result does not indicate any reluctance to continue flying.

All pilots expressed the importance of active support and personal commitment from others in their immediate surroundings, notably their flying colleagues. Professional assessment, counselling and supportive follow-up by flight medical officers and experienced flight psychologists, was considered to be very helpful, but a positive and supportive squadron attitude to accident survivors was regarded as the most fundamentally important factor in the complete rehabilitation of the pilot.

## INTRODUCTION

Studies concerning the emotional reactions following ejection are sparse. Fowlie and Aveline (1985) reported on 175 aircrew from the Royal Air Force who had survived ejection from an aircraft. They were asked about their emotional reactions after ejection, details of their rescue, and their rehabilitation after the accident. The authors, both psychiatrists, found that 40 per cent of the aircrew had developed prolonged emotional disturbances after the accident.

An aircraft accident is generally associated with thoughts or feelings of disaster, and is often a fearful and dramatic experience. For some, the accident may create new perspectives upon life of their professional situation, but others may be almost unaffected by the accident. It's difficult to measure emotional reactions. However, even though there are disadvantages, retrospectively, methodical questioning may provide a valid insight into the nature of human reactions to specific psychological experiences.

As a part of a flight safety program in the Swedish Air Force (SAF), the present study was initiated in 1986 by Kristina Pollack, chief psychologist in the Flight Safety Inspectorate Department of SWEDISH AIR FORCE STAFF and Hans Fries, LtCol and flight surgeon in the Flight Training School, F 5, SAF.

The objectives of this study were to obtain data regarding:

1) the emotional experiences of pilots who have escaped from incapacitated aircraft or survived chrashes

2) the thoughts and feelings associated with the accident

3) the attitudes to further flight service

4) the thoughts of the rehabilitation procedure after the accident and the pilots general ideas of how rehabilitation should be formed in the best way in the future.

## METHOD

The subjects in the study were 40 pilots who had been involved in aircraft accidents during the period 1978-85. 34 were members of SAF and 6 were helicopter pilots from the Swedish Army. Twenty-nine of the SAF pilots had ejected from high-performance aircraft, such as the Saab Draken or Viggen fighters.

All pilots were sent a comprehensive and detailed questionnaire to obtain objective and subjective information on a wide range of factors regarding the accident, and the subsequent reactions of the pilot to the event. Completed questionnaires were returned by all 40 pilots. Questions concerned the pilots flying status before the accident and the characteristics of the escape situation, such as time for decision, aircraft attitude, G-load, the ejection sequence, etc....

Data concerning the crash environment and the rescue were collected, e.g geographic location, temperature, time length of the rescue procedure, body injuries, etc. Furthermore questions concerning psychological/emotional reactions were asked, including attitudes before returning to flight service. Free comments were encouraged in responses to questions regarding desired psychological support and rehabilitation after an accident. All pilots were asked to rate their emotional experiences on the Life-Change Scale (LCS) devised by Holmes and Rahe. This scale is statistically derived from subjective ratings of the stress induced by emotional experiences of common induced life-events in a normal population. The 100 point scale ranges from minor life-events such as changes in habits of food to major life changes such as the death of a spouse (rated 100).

Answers concerning (1) the characteristics of the ejection -E- (2) the rescue procedure -R- (3) the medical/physical injuries -M- and (4) the psychological emotional reactions -P- were then categorized from 1 - 3. A "normal" case was represented by "1", minor disturbances by "2" and major deviations from "normality" by "3".

An ejection under controlled circumstances with time for planning, a stable flight configuration and little G-load was for instance categorized as "1". A rescue where a physically unharmed pilot was picked up with little time delay was regarded as "normal". If, however, the weather conditions were difficult or if there were complications in the rescue procedure the rating became "3". Minor physical injuries such as back pain were regarded as "normal", while fractures, burns, etc. were rated "3". Concerning the phychological/emotional reactions, longstanding difficulties of sleep or no motivation towards further flying were rated "3", etc...

## RESULTS

Within less than one month after their accidents two-thirds (65 per cent) of the SAF pilots had returned to flying (Figure 1). More than one-third of the total (38 per cent) returned to flying within one week after their accident. Of the group of pilots, who took longer than one month to return to flying half declared that they had been mentally prepared to return earlier, but were unable to do so due to physical injuries. Nine per cent of the total (3 pilots) were not allowed to return to flying for a variety of reasons unrelated to their psychological reactions to their accidents. One of these, however, did recommence flying after a two-year break.

Most of the pilots regarded the accident as a powerful emotional experience. One of them wrote: "The accident was a fantastic experience which I would not live without, but I would never like to experience again". One of the pilots drew an analogy with "a successful outcome of a fight with a bear".

None of the pilots expressed any reluctance to return to their professional military flying. In fact, the well-functioning rescue system seemed to reinforce a new feeling of safety.

In short terms, the majority of the SAF pilots expressed that they did somewhat change their attitudes to flying and to regulations. The awareness of the risks grew stronger. Immediately after the return to flying they flew with greater safety margins. After a while, however, most had returned to their pre-accident flying performance habits.

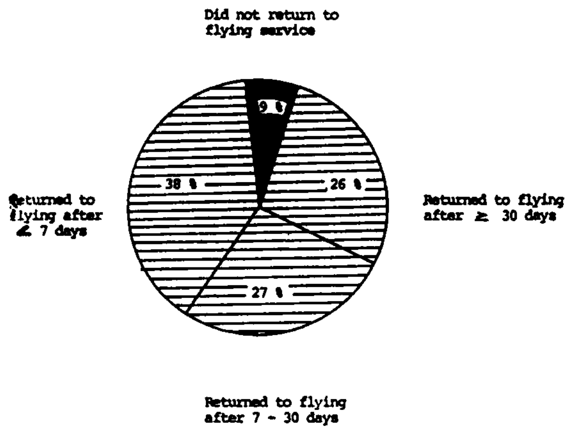



FIGURE 1 Timelength to return to flying for 34 SAF pilots

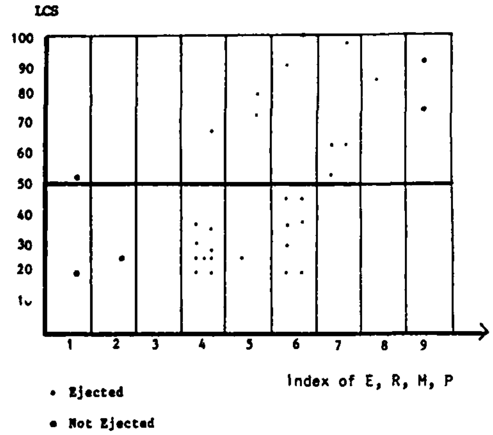



FIGURE 2 Distribution on the Life-Change Scale and categorized index of factors E, R, M, P

Some of the pilots stated that the accident had given them a new dimension to life, typified by expressions such as "enjoy it, as long as you have it".

Most of the pilots emphasize the importance of the "immediate surroundings" to take part and be engaged in "the human being of the pilot". Even though assistance from a flight surgeon and/or a flight psychologist could be of great value, the pilots considered the squadron attitude to be of the utmost importance to them in coping successfully with the psychological effects of the accident.

In figure 2 the distribution of the rating on the Life-Change Scale and the categorized index of the factors E, R, M, P is presented. 25 % of the answers were classified out of "normal" in which the ejection procedure and the medical/physical injuries were the main-represented factors.

More than half of the SAF pilots rated their emotional reactions below the point of 50 on the Life-Change Scale, but the deviation is considerable great.

## CONLUSION

Due to the complex nature of the problem, it is not possible to predict how an individual pilot will react after an aircraft accident. Among the pilots in the study the psychological reactions after their accidents may not be considered remarkable.

The pilots are basicly in their personality well prepared to meet critical situations or trauma. The majority returned to flying service after an interval of less than one month. It should be added that military pilots are generally highly motivated towards a demanding profession which they have chosen of their own free will. They are trained to work under stress, and are familiar with frequent exposure to high-risk situations. Consequently in their daily professional life they are often reminded of the serious consequenses of a single mistake. In short, they are mentally well-prepared to cope with the risks inherent in their operational performance - such as the risks of an accident.

There is also an inherent strength in a clearly defined homogeneous group such as a squadron - all members being men, similar in age and personality.

But there is another side of the coin. Some personal qualities will be more pointed out and strengthened and divergencies of personal disturbancies will be easily noted. A special life style and attitude in a group like this will more or less be developed to the expected performance of the pilot. Expected from people in general and not at least by themselves. Consequently, as a member in this group you will unconsciously performe in a way as close as possible the expected picture of the Pilot.

But the pilot is of course a human being like the rest of us. Even he will meet the crisis of life. Even he will live periodically under more or less stressful situations.

In the professional profile of the pilot honesty and trust to each other in the group is a fundamental condition for their mission. But you can also find a tendency not to show any affectness of emotional influence of so-called banal life-situations.

Most of the pilots inform, however, if they are in trouble or give a sign in a way, but even if the "immediate surrounding" has made their observations, they don´t interfere of personal considerations. After an aircraft accident these mechanisms may induce a denial or suppression of "normal" feelings, which delays a successful rehabilitaition. Thus, it is of great importance that harmful denial or suppression of feelings are recognized and dealt with by both the pilot and others close to him. The study confirms the importance of the personal care from collegues. Accordingly, it´s the ability of the collegues in the squadron to reach him - the pilot - that will be of greatest importance. It´s among them he wants to be accepted as the pilot he is and was, thereby not at least be accepted by himself.

Undoubtedly, some pilots would have liked a more personal and active commitment from significant others that had been forthcoming after their accident. Thus, the primary psychosocial milieu seems to be fundamental to successful rehabilitation.

As well as the causes of the accident are complex, there is no possibility to predict any general reaction of pilots after accidents. Every pilot must get his personal treatment after an accident.

The pilot´s basic psychological condition and situational condition, such as his position in the career, his life-situation, etc, as well as the accident´s divergency from a "normal procedure" are, however, factors to take in consideration. The question of guilt may influence the reaction as well as the consequences of the accident, for instance the possibility to return to flight service, etc.

The main objective of the flight psychologist and the flight surgeon should be to make the pilots more aware and able to understand their own feelings and behaviour and thereby give them the opportunity to cope in a more conscious, direct and successful way for flight safety from a human factor aspect.

## REFERENCE

Fowlie, D.G. & Aveline, M.O.:
"The emotional consequences of ejection, rescue and rehabilitation in RAF Aircrew"
BRITISH JOURNAL OF PSYCHIATRY
146: 609-613, 1985

Holmes, T.H. & Rahe, R.H.:
"The social readjustment rating scale"
J PSYCHOSOM. RES.
11: 213-218, 1967

# THE DESCENT FROM THE OLIMPUS
# THE EFFECT OF ACCIDENTS ON AIRCREW SURVIVORS

by

Lt. Col. Dr. I. Barnea
Chief of Psychology and Psychiatry
D. Efron Lurie M.A.
Head of Psychological Research.
Israel Defence Force, Air Force Aeromedical Center
Tel - Hashomer

## ABSTRACT

Limited research has been done studying the effects of flying accidents on surviving aircrews (pilots and navigators). In general, it consists of case studies conducted by squadron medical staff focusing on serious problems developed by the pilot and/or his family as a result of the accident.

The most significant finding of this research has been that the return to flying following an accident in flight is not as simple as it may appear. To consider an accident as part of the training routine and to expect return to normal activities as if nothing out of the ordinary has occurred is to ignore the legitimate emotional reactions of the aircrew and the possible effects of those reactions on the development of physical and psychological symptoms including deterioration of flying performance, motivation and commitment.

Return to flying with "hidden", untreated symptoms is likely to aggravate them, prolong their resolution or even make them irreversible resulting in the loss of flight personnel.

The present study has two main objectives. The first, to improve our understanding of the relationship between a flying accident, the resulting emotional reactions and their effect on performance levels. The second, to develop an effective intervention procedure to enable the accident survivors to return quickly to preaccident functioning, both emotionally and professionally.

To reach this end, several aircrew survivors of serious flying accidents in the past five years participated in a research conducted by a psychologist using a structured interview especially developed for the present study.

Results of the interview showed that 44 percent of the survivors reported a decrease in their flight performance and feeling toward flying than prior to the accident. This change lasted a few months. The most vulnerable population to accident effects was the young pilots. Following the accident the myth that "this can't happen to me" was broken and aircrew survivors became more aware of the risks of flying putting more emphasis on flight safety and professionalism. Most of the survivors (91 percent) felt they needed to talk with friends and family members for the purpose of "emotional ventilation" as well as with superiors and colleagues for the purpose of strengthening their flying skills. A certain percentage of survivors felt they needed help from a psychologist who was considered by them to be neutral being associated neither with the squadron nor with the committee investigating the accident.

In light of the findings generated from the interviews an intervention procedure is proposed to help survivors return quickly to their preaccident spirit and performance.

## INTRODUCTION

Despite the fact that many resources have been allocated to the development of means for surviving flying accidents (especially in fighter aircraft) and in the provision of medical services to aircrew following such accidents, only scant attention has been paid to the short term and long term effects of flying accidents on their survivors (Fowlie and Aveline, 1).

Aircrew members are expected to perform efficiently in the face of daily dangers. An accident is an event which can cause him to reconsider those dangers associated with flying and give rise to difficult emotional reactions as well as deteriorating motivation and flying performance. Until recently, these reactions have not been given sufficient recognition. Instead, survivors are expected to return to flying the very moment their medical status allows and to ignore or block all expressions of feelings. The message to ignore or block feelings comes from a variety of sources beginning with the squadron which considers the accident and coping with it to be part of the requirements of the job and ending with fellow flight personnel and family members. As a result, most survivors of flying accidents indeed return to flying soon after the accident, some with clear success (Fowlie and Aveline, 1).

Jones (2,3) has found that many pilots cope with the daily dangers of flying by using a complex combination of defense mechanisms comprised mainly of denial, repression, rationalization, intellectualization and humor. Moreover, he concludes that the use of denial is not only characteristic of pilots but also vital to their combat performance, being viewed, therefore, as an adaptive mechanism. Self confidence and the absolute faith in their continued ability to rely on themselves is also characteristic of successful pilots. However, it is these same mechanisms and characteristics which, in the face of dangerous events that arouse doubts concerning their control or their ability to predict the reoccurrence of uncontrollable events, cause stress and may even become the source of sicknesses and/or of emotional difficulties, especially after accidents.

Work undertaken by medical staff members of squadron in collaboration with teams of psychiatrists who treated pilots suffering from emotional reactions following accidents has made it clear that the return to flying of air personnel after an accident is more problematic than may be assumed from superficial examination ( Fowlie and Aveline, 1). Even though enough time has passed and the event has been filed and made a part of history, the accident still upsets the survivors both on ground and in the air taking such forms as phobias, situational adjustment reactions, oversensitivity to vulnerability and somatization. These reactions can influence various flying situations (Jones, 3) resulting in a deterioration of flight performance, motivation and commitment.

Popplow (4) attributed to this phenomenon the title of "post accident anxiety syndrome". According to him this phenomenon exists, but is not recognized, since it is a very sensitive subject which pilots are not willing to discuss openly preferring to divert conversation from it and their feelings to concerns about the families of the killed or injured, to technical evaluations of the event, the weather and the skills of the pilots who experienced the accident. The general message is "this can't happen to me".

Aitken (5) states that all aircrews experience anxiety when they are exposed to the threat of danger. Some even perform less well especially in a squadron in which the most recent accidents resulted in physical injury. Trimbel (in 1) asserts that following life threatening events, neurotic symptoms are expected even in the absence of prior personality vulnerability.

Zeller (6) studied the emotional reactions of 200 United States Air Force aircrew who had survived combat accidents mostly by ejection. Thirty three percent showed clear emotional consequences including aimless and inefficient behavior, panic, time distortion and transient psychotic states.

Fowlie and Aveline (1) examined 175 pilots in the Royal Air Force who bailed out. They found that 40% of the pilots suffered from emotional disturbances such as fear, anger, worry, disgust, stress and change in motivation. Of them 28% returned to flying although they continued to suffer anger, fear and apprehensiveness which surely affected their performance and motivation. Thirty one percent of the subjects in this study indicated that empathic counseling during rehabilitation had been of crucial importance to their emotional recovery.

From the limited literature on this subject it can be seen that the return to flying after an accident is not problem free. Thirty to forty percent of accident survivors have various complications. Not only is the problem not handled, it is in most cases concealed. Referrals to counseling services (medical and psychological) come too late, usually when the symptoms worsen to such a degree as to be out of the control of the crew member and resulting in temporary or permanent grounding. The problem becomes more serious if we take into account the fact that a great number of aircrews experience some kind of accident during their service (both in peace time and war time) while those remaining experience them as bystanders. If accidents affect 30-40 percent of aircrews, then that constitutes a problem requiring an appropriate and effective solution.

The object of the present research is to improve our understanding of the relationship between the accident, it's possible consequences and the needs of the survivors. Improved understanding will aid in the building of a program of professional intervention based on principles of prevention and treatment which will care for aircrew survivors following an accident. In this way "secondary casualties" can be prevented which could be the result of faulty or too late treatment thereby lessening the damage incurred from the accident.

## METHOD

### SUBJECTS

A classified number of aircrew accident survivors participated in the study,(their number were sufficient for statistical analysis). All of them experienced very serious flying accidents within the past five years. Two thirds were pilots and one third navigators. Half of the survivors examined (the "veterans") had served in the Air Force more than seven years prior to the accident while the other half (the "youngsters") served for seven years or less prior to the accident.

### INTERVIEW PROCEDURE

According to Israel Air Force procedure, all aircrew members who have undergone a serious accident are required to appear immediately at the Aeromedical Center for a meeting with a psychologist. The psychologist represents a neutral body which is not associated either with the accident interrogation committee or with the squadron command. The meeting is required for returning to flying. The main goal of the meeting is to reconstruct the event, work it through and evaluate its immediate influence. As a result recommendations regarding the return to flying are made.

On the basis of material collected from such meetings in the past number of years and on the basis of the questions put by the present research, a structured interview was designed to collect data on the following subjects: Description of the accident and its context; feelings and thoughts regarding the accident and toward flying as a result of the accident; flight performance, motivation and commitment to flying following the accident: the influence of the accident investigation on the aircrew member: advantages arising from the accident: and the survivors' recommendations for assisting their professional and emotional rehabilitation including the role of the squadron commander in such rehabilitation.

### PROCEDURE

Each subject was invited to meet with a psychologist who explained the purpose of the study and conducted the structured interview built especially for the research.

## RESULTS

According to the analysis of the structured interviews, the results were organized into three sections.

### THE MAJOR FEELINGS STATES RESULTING FROM FLIGHT ACCIDENTS

Clinical experience indicated that an accident can be defined by three time periods: Feelings during the accident, feeling during the first 48 hours after the accident and feelings occurring during the first six months after the accident. In the present research the most important feelings were mapped in accord to each of these three time periods.

Principal feelings characteristic to the accident in progress:

| The Feeling | Percent of survivors experiencing |
|---|---|
| Surprise | 94% |
| Sharp, clear thoughts | 79% |
| Fear | 64% |
| Thought acceleration | 64% |
| Narrow, sharp, focused attention | 64% |
| Time distortion (slow motion) | 51.5% |

Principal feelings characteristic of the first 48 hours after the accident:

| The Feeling | Percent of survivors experiencing |
|---|---|
| Recurrent pictures of the event | 72% |
| Relief | 70% |
| Feeling that one's life was given as a gift. | 61% |
| Joy | 55% |
| Guilt (I could have done differently) | 52% |

Principal feelings occurring within the first six months of the accident:

| The Feeling | Percent of survivors experiencing |
|---|---|
| Recurring images of the event | 73% |
| Fear of flying | 52% |
| Lack of confidence during flying in a similar situation | 48% |

The results of this mapping of emotional reactions are of significant importance in the building of an intervention strategy which will be discussed later.

### THE EFFECT OF THE ACCIDENT ON FLIGHT PERFORMANCE AND FEELINGS TOWARD FLYING

Twenty four percent of the air crews had thoughts, although minor, of not returning to flying as a result of the accident.

Fifty one percent reported that their feeling about flying had changed. These were expressed as loss of confidence, tension, apprehension and worry in situations similar to that when the accident occurred. Forty four percent of the survivors reported a deterioration in their flight performance. These changes continued from several days after the accident to several months. (A number of survivors reported that the above symptoms continue even today, some several years after the accident.)

These findings confirm the basic premise that there is a relationship between the accident and changes in feelings and performance of aircrew survivors in a large portion of them (at least 44 percent). This strengthens the need for professional intervention after the accident in order to enable the crew members to return to preaccident levels of performance and emotional well being.

An additional finding is related to the delineation of two distinctly different aircrew populations, the veterans and the youngsters. Fifty percent of the youngsters reported a deterioration in flight performance after the accident in comparison with 31 percent of the veterans. Sixty five percent of the youngsters reported a change in their feelings toward flying in comparison to 36 percent of the veterans. This finding is important when considering intervention strategies in that the two populations differ in their degree of vulnerability.

## PERSONAL LESSONS LEARNED FROM THE ACCIDENT AND SURVIVORS' SUGGESTIONS FOR RECOVERY

According to the survivors three interdependent conclusions were drawn;

The myth that "this can't happen to me" was broken for twenty two percent of the survivors resulting in heightened awareness of the risks of flying and of their own vulnerability to injury which had been denied until the accident . As a result, they felt the need for reevaluation of their approach to flying.

As a result of this reevaluation, they concluded that they need to improve their level of expertise, to collect more information and to practice handling more and varied types of emergency situations based on the principle of anticipation ("provide the cure before the blow"). Survivors reported that their confidence in the ejection equipment and in training for emergencies increased. Those feeling are shared by 57 percent of the survivors.

As a result of the loss of the myth of invulnerability, their attitude toward flying after the accident is more mature and prudent. Twenty percent of the survivors reported a decrease in attitudes of indifference and impetuousness at least for some time after the accident.

Ninety one percent of the survivors felt the need to talk for the purpose of ventilating feelings with friends and family and for the purpose of receiving reinforcement for their professional ability from their commander and fellow crew members. More specifically there is need to recreate the event in detail analyzing and evaluating their professional performance as it unfolded. It is very important to the survivor that his professional competence displayed during the accident be affirmed even in an open forum. A certain percentage of survivors expressed an interest in meeting with a psychologist for the purpose of clarifying their specific problems with a neutral person not associated with the accident investigation committee or the squadron.

Most of the survivors suggestions for recovery covered three major areas:

### Professional rehabilitation:

It was recommended that the crew member return to flying after a thorough and in depth debriefing of the accident. Some of the survivors suggested a gradual return to missions according to level of difficulty, under supervision or accompanied by a stronger, more experienced crew member. The purpose of this is to assist the survivor in rebuilding his confidence.

### Emotional rehabilitation:

It was recommended that the survivor not hold his feelings in or hide them but to share them with family members, friends, colleagues, crew members who had undergone accidents in the past and superior officers. The purpose is to make legitimate the various emotions the survivor experiences and to ensure that the survivor doesn't feel exceptional in what he undergoes but instead realizes that his reactions are typical of all the reactions that result from flying accidents.

The role of the squadron commander in assisting the process of rehabilitation:

According to the survivors recommendations, the role of the commander is four fold.

Thorough debriefing of the accident

The commander must analyze the event with the survivor for good and for bad, and to summarize the matter in a clear way without leaving room for imagination or misunderstanding. The discussion must be open enough to deal with anger and accusations.

Support

The survivors expect to receive support from their commander especially when they acted correctly. They expect him to express his confidence in their flying ability. "When a pilot is allowed to fly after an accident, it is the greatest support he can receive. This means that he can be relied upon." If the survivor is responsible for the accident, he expects the commander to defend him against attack from outside the squadron.

Conversations

The commander must be available to conduct informal conversations with the survivor in order to promote ventilation, soothing, affirming the survivor's status as a squadron member and dealing with fears that may have arisen.

Follow up

The commander is expected track the survivors for some time in a discrete manner monitoring their feelings and functioning.

## DISCUSSION

Aircrew members are very likely to experience flying accidents in the course of their flying careers, either first hand or as a witness of the accidents of others. Despite the scarcity of research conducted in this area, it can be concluded that accidents adversely affect at least 40 percent of the survivors resulting in the development of emotional problems and the deterioration of flying performance.

The results of the present study support these conclusions: At least 44 percent of the aircrews who survived flying accidents reported negative changes in two major areas of functioning: flight performance and feelings before and during flights. These changes generally persisted for several months after the accident. Moreover, it was found that the younger pilots were most at risk to develop post accident difficulties. This result is of great importance when considering the building of intervention programs.

Aircrews are a unique population in that they exist under conditions of perpetual competition which demands the suppression of any and all signs of weakness. The accident and its effects place a two-fold pressure on the survivors for not only do they have to live with the fact of the accident, they also interpret their emotional reactions as illegitimate and a sign of illness. As one crew member said, "Not only did I have an accident, I now am also crazy". This process of delegitimation results in the concealing of these feelings at the cost of a decrease in flight performance and motivation. If not resolved the cost can increase in the development of psychological and psychosomatic symptoms and ultimately the cessation of flying altogether. The present research outlined a map of the major emotional reactions experienced during the accident, immediately after (48 hours) and much later (after 6 months). The mapping of feelings, then, is intended to be used in the creation of a counseling aid which, by delineating and defining those emotional sequelae as legitimate, will facilitate the opening of communication between the survivor and those close to him. Instead of viewing his feelings as a sign of weakness and illness, something to be concealed, they will be viewed as signifying a relevant problem in need of solution.

Methodologically, it was concluded that because in the population of aircrew personnel the use of denial is so all pervasive and legitimate, there is a need to develop special questionnaires to minimize hiding behind "healthy" or neutrally non-revealing answers. Moreover, it is preferable that the interviewer is not related in any way to the accident investigation committee or squadron medical staff.

On the basis of the mapping of emotions and the recommendations of the aircrew survivors, we propose an intervention program that will be implemented immediately after the accident thus treating the problem from the standpoint of prevention. The program is aimed to deal with three levels of the problem: The individual, the command and the squadron. It's major points are as follows:

Immediately after an accident the survivors will be given a short and informative paper delineating the most common emotional reactions to air accidents and what can be expected in the short and long term. This will be available to them at all times.

The commander will be aware of the need for survivors to undergo a thorough, in depth debriefing, to receive support and to be available for informal conversations for the ensuing months.

Shortly after the accident the entire squadron will convene and conduct a group discussion with five goals in mind:

A discussion of the events of the accident and how best to handle similar situations. The purpose is to provide information to the squadron staff.

Public support for the survivor providing he performed as expected.

Public disclosure of experiences and emotional reactions undergone by squadron members who also were in accidents. The purpose of this is to facilitate and lighten the emotional coping of the survivor of the present accident.

All information is to be provided and all questions asked and answered in order to minimize and prevent bothering the survivor with repeated queries over time.

The group discussion will enable those squadron members who were witnesses or bystanders to express their apprehensions and emotions.

A psychologist will be in contact with the survivor from forty eight hours after the accident, a procedure already instituted by the Israel Air Force and recommended to be continued. The purpose of the first meeting is to evaluate the mental status of the survivor and his capacity to return to flying. The psychologist will act as a consultant to the squadron commander who will be the main agent in the rehabilitation of the survivor. Therapeutic intervention will take place only as needed. A meeting between the psychologist and the survivor which will take place soon after completion of the accident investigation will signify closure of the matter.

For the entire time any intervention and it's depth will be determined by the type of accident, the damage incurred, the loss of life involved, the seniority of the crew members involved and the degree of responsibility or guilt over the accident.

## REFERENCES


1. Fowlie, D. G. & Avline, M. O. (1985) The emotional consequences of ejection, rescue and rehabilitation in Royal Air Force aircrew. British Journal of Psychiatry, 146, 609 - 613.

2. Jones, D. R. (1982) Emotional reactions to military aircraft accidents. Aviation, Space and Environmental Medicine, June, 595 - 598.

3. Jones, D. R. (1986) Flying and danger, joy and fear. Aviation, Space and Environmental Medicine, February.

4. Popplow, J. R. (1984) After yhe Fire Ball, Aviation, Space and Environmental Medicine, April.

5. Aitken, R. C. B. (1972) A study of anxiety assessment in aircrew. British Journal of Social and Clinical Psychology, 2, 44 - 51.

6. Zeller, A. F. (1973) Psychological aspects of aircrews involved in escape and evasion activities. Aerospace Medicine, 44, 956 - 960.

## DISCUSSION

SIOMOPOULOS: I have a question which relates to both papers. Did you look into pilots who had made an error, that is, the accident was their fault, versus pilots who were not responsible because of something mechanical? I have in mind an accident in which the aircraft caught fire, the pilot remained cool and did what he had to do. Also he applied power in order to make sure that he avoided an inhabited area. This pilot's feelings on the effects of the accident may not be the same as if he did something stupid, like putting the aeroplane out of control and having to eject. That is one question.

The second question raises the question of denial. What you are trying to do is against the acceptance of denial as a useful mechanism. If you look into your paper you'll see that the younger pilots were using denial to a very large degree and werė helped by whatever enthusiasm they had. The pilots who were more experienced used less denial and were really more aware of all the dangers of the job. You have different effects in the two groups. The idea of accepting denial as a useful mechanism is a concept I would doubt rather seriously. I would like the pilot to know what he is doing, to know the danger to begin with. When he is young he does not listen to what you tell him about the danger, but at least he should rely on denial as little as possible. That is my experience with our pilots.

VAN DEN BIGGELAAR: My question is more or less related to the previous one. In your research you had a group of subjects who had all experienced accidents or incidents. Did all these incidents or accidents occur in peacetime or were some of the accidents in combat situations? In the latter there is going to be another mechanism. I would like your opinion on the possible differences in mechanisms following accidents that occurred in peacetime and in combat.

BARNEA: All the accidents were during peacetime and not during war. Now to the first question, I think the whole question of guilt was very essential. Pilots who felt responsible or guilty for their accident had much stronger reactions for a much longer period.

POLLACK: I would like to add to this that we also found feelings of guilt in pilots involved in technically related accidents. They always think they could have done something else during these stress situations; guilt is found in both types of accident. Concerning the study we did, we were looking at both the human factor related accidents and the technical ones.

PSIMENOS: I would like you to comment on your description of the descent from Olympus. This means that there was an ascent to Olympus; in order to descend from Olympus you have first to ascend it. This means, perhaps, that the pilot population has a feeling of superiority and that they are inviolable. The accident proves the opposite, so the pilot has not only guilt but also disillusionment. Is this a matter of training which produces a general spirit in Israeli pilots such that they feel they are superior and not prone to accidents?

BARNEA: I have heard also from other people this morning that the basic 'macho' personality that we were talking about exists both in the Israeli Air Force and all other Air Forces. I think it is part of the profession. When we are trying to talk to them about safety they do not want to listen. They do not want to hear, probably, because of defence mechanisms. They reject everything that is connected with vulnerability or with the fact that something might happen to them, especially the younger aircrew population. They do not want personal failures to be exposed to other pilots in the squadron etc. We are trying, right now, to work much more closely with the pilots on the squadrons, for instance, with issues on how to avoid failure. Israeli pilots do not know how to deal with failures.

Is failure a part of training? Do you have to fail in order to learn something? What happens is that aircrew, who have been involved in an accident and something was broken in their defence mechanism, are more in touch with their emotions and with the risks. Some say they do not enjoy flying anymore. They do not say it in public, but they will say it in private. Some of them say it is very helpful and they become better pilots. When they talk about their feelings and attitudes their mates do not want to hear what they have to say even if it is very emotionally expressed. So here there is a collective defence mechanism of the group of fighter pilots. I have questions about it because when we are talking about crew co-ordination and things like that it is very clear that they have to be aware and you can show them that it can be helpful to them. It is very much, in a way, task orientated, but sometimes when you are talking about safety and judgement they do not want to deal with it. They do not prepare themselves for an event that might happen and they do not learn from their friends who were involved in accidents. So this is a problem that we are dealing with right now, and it is difficult. You have, on the one hand, to appreciate the mechanisms. On the other hand, I think that we as psychologists can add something that can help them.

STEVENSON: Dr David Jones described the "I should have died" syndrome that is seen in some people who are involved in accidents in which they almost get themselves killed. They develop a behavioural pattern later on of trying to complete the act by engaging in risk taking behaviour. They take up motorcycle riding, sport parachuting, or engage in more dangerous types of flying. I wonder if either of you have noticed any of this kind of behaviour which is the opposite of what you describe in people who have had mishaps?

BARNEA: We had two accidents in which the pilots had been involved in previous accidents. When they experienced a spin during air to air combat they didn't want to eject, probably because of the accident they had in the past. This means that, somehow, they did not debrief the first accident in a way that was productive for them. These two pilots were highly competitive people so it may be that they didn't eject because of the previous accident; but this is very rare.

AUFFRET: We have had five papers, five different appproaches, coming from five different countries, but I have only four comments to make. Firstly, I think it has been said by one of the speakers that it is too simplistic to separate all the accidents caused by human factors from those caused by external or mechanical failures. I think it is too simplistic because it tends to provide a too immediate and too simple explanation of the accident; I think the causal factors are much more complex than that. The second comment relates to the roles of the flight surgeon and psychologist. In my opinion, it is better to try to understand why the accident happened, so as to be able to advise higher command and develop good prevention, rather than making accusations or condemnations. The latter should be done by higher command or by a commission of enquiry. The third comment, I am talking as a former pilot and I have been a pilot for many, many years, is that I believe aircraft are becoming more and more easy to handle; very often it is the computer that will offset and repair an error of the pilot, I am talking only in terms of pilotting. On the other hand, the mission is becoming more and more difficult. My last comment is that I have been surprised, I must admit, to see the importance of the psychological follow-up after an incident or accident.

# PERSONALITY CHARACTERISTICS OF USAF PILOT CANDIDATES

by

Dr Frederick M.Siem
Personnel Research Psychologist
AFHRL/MOFA
Brooks AFB, TX 78235-5601, USA

## SUMMARY

To examine the utility of personality measures for enhancing current selection methods, a computerized personality inventory was adminsistered to a sample of United States Air Force pilot candidates prior to flying training. Analysis of the data suggested that two of five personality factors were associated with training outcomes, and that one of the measures added predictive utility to test scores currently used for pilot selection. Candidates who were self-confident and not dogmatic manifested higher graduation rates than pilot candidates who were either less self-confident or more dogmatic (less flexible in their values).

For a subset of respondents, performance scores were available for two phases of training. Examination of the data indicated that personality characteristics (depression, activity level) that did not differentiate training graduates from non-graduates were associated with better performance in two phases of flight training. In contrast, characteristics on which graduates and non-graduates differed (self-confidence, values flexibility) did not appear to be associated with performance scores during training.

The main conclusion drawn from this research is personality measures can contribute predictive utility to a pilot selection system over and beyond that displayed by currently operational aptitude measures. A second conclusion is that careful consideration must be made in the selection of both predictor and criterion variables in quantifying the relationships best suited for determining operational utility of personality measures.

## INTRODUCTION

Since World War I, psychological tests have been used for the selection of military aviators in the U. S. [1]. Although measures of personality characteristics were employed in World War I, the trend during and since World War II has been toward a greater reliance on tests of psychomotor coordination and cognitive skills [2]. This trend can be explained, in part, by the observation that the tests developed by the Army Air Forces in the 1940s laid the foundation for both the current operational paper-and-pencil test used by the U.S. Air Force (USAF) and for an experimental computer-administered test battery recently validated for pilot selection [3]. Over the years standard procedures have evolved for measuring 'stick-and-rudder' skills and cognitive abilities, such as information processing efficiency and spatial visualization skills. In contrast, for personality measures no such standardization exists. Research into pilot characteristics and their relationship to aviation performance has produced a substantial body of literature but has not clearly identified any single instrument for operational use as an aid to military pilot selection [4].

Several factors may account for this situation, such as the wide variety both in predictor and criterion measures used in different studies. Much of the earlier research, for example, focused on the relationship between clinical scales and pilot training measures [5], whereas more recent studies have employed scales developed for 'normal' populations [6, 7, 8]. Also, several studies have focused on differences between members of the aviation community and the general population [9, 10, 11], while others have examined the relationship between pilot characteristics and safety issues [12, 13, 14].

Another factor that may account for the lack of a cohesive set of findings concerns a statistical issue. Many studies have examined only simple linear relationships between a large number of intercorrelated personality measures with some criterion. One would expect that the results would vary somewhat from sample to sample, given the same set of measures, simply as a consequence of multicollinearity. Comparison of studies, then, using somewhat different instruments becomes all the more complicated, and rarely has the practice been to extract more global measures from the personality instruments employed.

Despite this lack of conclusive research findings regarding the utility of personality measures for pilot selection, a number of countries currently demonstrate at least an implicit acknowledgment of the importance of personality factors for military pilot selection, as manifested by procedures that involve clinical screening and interview processes [15]. Most of those countries can utilize such techniques, requiring intensive screening by highly trained personnel, because of centralized selection procedures. Pilot candidates for the U.S. Air Force, in contrast, are drawn from three sources with somewhat different selection procedures: the Air Force Academy (AFA), the Air Force Reserve Officer Training Corps (AFROTC), and the Air Force Officer

Training School (OTS). Applicants for all three sources are dispersed across a broad geographical area and assignments to pilot training requirements are determined by a number of different selection boards; hence, the need exists for standardized instruments and screening procedures that can be employed in a decentralized fashion.

In response to on-going Air Force concerns about potential techniques to improve its pilot selection procedures, the present research was designed to evaluate the potential of standardized measures of attitude and personality for reducing attrition and enhancing the quality of pilot trainees. The research was designed both to build upon previous studies and to address specific limitations of the previous research for Air Force pilot selection.

## METHOD

### Sample

The sample consisted of 509 candidates from USAF Undergraduate Pilot Training (UPT). Most of the sample was male (99%; N = 503) and the average age of the pilot candidates was 23.8 years.

### Procedure

Pilot applicants were selected into the Air Force on the basis of, in part, achieving minimum scores on the Air Force Officer Qualifying Test (AFOQT). The AFOQT is a paper-and-pencil aptitude battery that includes sixteen subtests which are combined into five composites, of which two measure aptitude for pilot training: a Pilot composite and a Navigator-technical composite [16]. Prior to entry into flying training, the sample was administered the Automated Aircrew Personality Profiler (AAPP), the instrument used in this study to measure personality characteristics.

The Undergraduate Pilot Training (UPT) program in which respondents participated lasted 49 weeks. The program was composed of two phases of flight training, one in T-37 aircraft and the other in a T-38 aircraft. The T-37 was a subsonic, low performance jet training aircraft, whereas the T-38 was a high-performance, supersonic jet aircraft. Each phase of flying training was further subdivided into three subphases. At the end of each subphase, a graded check flight was performed on which the student pilot was given a percentage score, a ratio of performance points achieved to points possible. Thus, the six check flight scores could vary from 0 to 100%.

Candidates could be eliminated from training at any point in the program for a variety of reasons. The most common reasons for elimination were flying training deficiency (FTD), medical problems (such as air-sickness), Manifestation of Apprehension (MOA) and self-initiated elimination (SIE), with a smaller number of candidates eliminated for poor academic performance or military demeanor.

Candidates remaining in the program at the forty-second week of UPT were considered for follow-on instruction by an Advanced Training Recommendation Board (ATRB) in either a fighter-attack-reconnaissance training track (FAR) or a tanker-transport-bomber training track (TTB). The ATRB was a panel of T-38 Instructor Pilots who determined whether candidates were better suited for fast jet operations (the FAR track) or for flying TTB aircraft. Actual aircraft assignment was a function of three factors: the ATRB outcome, individual preference, and aircraft availability. Information about graduation or reason for elimination, as well as the ATRB recommendation, was collected at the end of UPT.

The number of respondents having data for each of the predictor and criterion measures is shown in Table 1. For some of the sample tested on the personality measure neither training outcome data nor AFOQT scores were available. For a subset of training graduates, the ATRB measure was also unavailable.

Table 1. Sample Size for Predictor and Criterion Measures

| Measure | Sample Size |
|---|---|
| Automated Aircrew Personality Profiler | 509 |
| Air Force Officer Qualifying Test | 292 |
| Check Flight Grades | 144 |
| Undergraduate Pilot Training Pass/Fail | 325 |
| Fighter/Non-fighter Recommendation | 224 |

Instrument

The Automated Aircrew Personality Profiler (AAPP) consisted of 202 items representing scales from several instruments: the Minnesota Multiphasic Personaltiy Inventory (MMPI), one of the more commonly used diagnostic tools in clinical practice [17]; the State-Trait Anxiety Inventory [18]; the Personal Orientation Inventory [19], an instrument designed to assess an individual's aptitude for self-actualization; the Interpersonal Behavior Scale [20], which measures assertive and aggressive tendencies; and the Jenkins Activity Survey [21], designed to measure personality factors associated with chronic heart disease.

Analyses

An exploratory factor analysis was conducted on the AAPP personality scales in order to reduce the number of predictor measures. Next regression analyses were used to examine the utility of personality measures for predicting pass/fail outcome, both alone and in combination with the selection measures used in the current system. Personality differences between training graduates and non-graduates were also examined with regression analyses. Finally, relationships between personality measures and performance at different phases of training were compared using correlational analyses.

## RESULTS

Factor Analyis

As a result of analyses reported in detail elsewhere [22], nine of the original twenty-five scales employed in the AAPP were dropped from further consideration either for exhibiting poor internal consistency or for low zero-order validities with training

Table 2. Scale Composition of Five Factors (N = 509)

| Scale | Factor I | II | III | IV | V |
|---|---|---|---|---|---|
| Manifest Hostility | 69 | | | | |
| Need for Affection | -68 | | | | |
| Naivete | -92 | | | | |
| Distrust | 68 | | | | |
| Ego Inflation | 35 | | | | |
| Frankness* | | 36 | | | |
| Denial of Social Anxiety | | 84 | | | |
| Social Imperturb-ability | | 99 | | | |
| Imperturbability | | 55 | | | |
| Acceptance of Aggression** | | | 66 | | |
| Values Flexibility** | | | 64 | | |
| Poignancy | | | | 57 | |
| Brooding | | | | 69 | |
| Amorality **** | | | | | |
| Hypomania | | | | | 67 |
| Hard Driving*** | | | | | 35 |

Note: Only factor loadings with absolute value > .30 are shown. All scales from MMPI unless otherwise indicated. Factor I = Hostility; Factor II = Self-confidence; Factor III = Values Flexibility; Factor IV = Depression; Factor V = Activity Level.

* Scale adapted from Interpersonal Behavior Scale
** Scale adapted from Personal Orientation Inventory
*** Scale adapted from Jenkins Activity Survey
**** Scale did not load substantially on any factor

measures. The remaining sixteen scale scores from the AAPP were factor analyzed using principal factoring with iteration and oblique rotation. Five factors emerged with eigenvalues greater than 1.0 on which all but one scale (Amorality) manifested loadings with an absolute value greater than .30 (see Table 2).

Based on the exploratory factor analysis, five personality scores were computed. Each of the five measures was derived from combining the raw scores for each of the component scales. Scales with negative factor loadings were reverse scored before being summed into the factor scores. From inspection of scale items, the factors were interpreted to be global measures of 1) hostility, 2) self-confidence, 3) values flexibility, 4) depression and 5) activity level.

## Relationship Between UPT Pass/Fail and Personality Factor Scores

The point-biserial correlations between the five personality factors and UPT training outcome (pass/fail) are shown in Table 3. As these data indicate, three of the scales were associated with UPT outcome. Note that the graduation rate was greater than 80%, so that the p/q split for this sample was .82/.18. Thus, the maximum correlation that could result from these data is .7 [23].

Table 3 also includes, for UPT graduates only, point-biserial correlations between the personality scores and the recommendation for fighter/non-fighter follow-on training. The ratio of fighter-recommended to non-fighter recommended candidates was .55 to .45 in this sample. None of the zero-order correlations approached significance.

Table 3. Correlations of Five Personality Factor Scores with UPT Outcome and Follow-on Training Recommendation

| Personality Factor Score | UPT Pass/Fail ($n$ = 325) | FAR/TTB ($n$ = 224) |
|---|---|---|
| Hostility | -.12* | .01 |
| Self-confidence | .13** | -.01 |
| Values Flexibility | .12* | .10 |
| Depression | -.10 | -.03 |
| Activity Level | -.02 | .00 |

Note: FAR = follow-on recommendation for fighter training, TTB = follow-on recommendation for tanker-transport-bomber training.

* $p < .05$ that variation from zero correlation is due to chance
** $p < .01$ that variation from zero correlation is due to chance

## Operational Utility

To examine the operational usefulness of the measures generated in this study, a multiple regression analysis was conducted. The criterion was a dichotomous variable representing training outcome (graduate/non-graduate). The predictor set consisted of scores for the Pilot and Navigator-Technical composites of the AFOQT, as well as scores for the five AAPP personality factors. The predictors were entered in two stages: at the first stage, the AFOQT scores were entered, and at the second stage the personality factor scores were added using the stepwise method, with the criterion for entry of new variables being a significance test level for incremental validity of .05. This analysis indicated that only the self-confidence measure added incremental validity to those test scores operationally used in the current system ($R$ change = .07, $F$ [1, 289] = 3.99, $p < .05$).

## Personality Differences Between Graduates and Different Types of Non-graduates

To explore whether UPT graduates exhibited different personality characteristics than the various types of non-graduates, a set of multiple regression analyses was conducted. For each analysis, the dependent variable was one of four binary vectors representing pilot training outcome: graduate vs. all types of non-graduates, medical eliminees vs. graduates and non-medical eliminees, self-initiated eliminees (SIE) vs. graduates and non-SIE eliminees, and eliminees due to Flying Training Deficiency (FTD) vs. graduates and non-FTD eliminees. For each regression the predictor set consisted of the five personality scores, which were entered in a stepwise fashion.

Based on this set of analyses, it appeared that graduates differed from non-graduates on both of the socially desirable personality characteristics, self-confidence and values flexibility ($\underline{R}$ = .17, $\underline{F}$[2, 322] = 4.82, $p < .01$). Note that the hostility measure, although significantly correlated with the pass/fail measure, did not contribute any additional utility to a model for differentiating graduates from non-graduates.

The self-confidence measure also differentiated medical eliminees from the rest of the sample ($r$ = .16, $\underline{F}$[1,323] = 8.39, $p < .01$), whereas the values flexibility measure distinguished flying training deficiency eliminees from the rest of the sample ($r$ = .12, $\underline{F}$[1, 323] = 4.91, $p < .05$). As the means in Table 4 indicate, UPT graduates reported more self-confidence ($\underline{M}$ = 20.38) than medical eliminees ($\underline{M}$ = 15.60). Graduates also reported higher scores on the values flexibility measure ($\underline{M}$ = 13.26) than did flying training deficiency eliminees ($\underline{M}$ = 12.30).

Table 4 Mean Personality Factor Score for UPT Graduates and Elimination Groups (N = 325)

| | | Eliminee Group | | |
|---|---|---|---|---|
| Personality Factor | UPT Grads ($n$ = 267) | FTD ($n$ = 34) | MED ($n$ = 10) | SIE ($n$ =14) |
| Self-confidence | | | | |
| Mean | $20.38_a$ | $18.85_a$ | $15.60_b$ | $20.64_a$ |
| SD | 4.53 | 4.77 | 6.20 | 5.33 |
| Values Flexibility | | | | |
| Mean | $13.26_a$ | $12.30_b$ | $13.90_{a,b}$ | $12.43_{a,b}$ |
| SD | 2.74 | 2.80 | 3.38 | 3.13 |

Note: MED = Medical Eliminee; FTD = Flying Training Deficiency Eliminee; SIE = Self-Initiated Eliminee. Means with a different subscript differ at the $p < .05$ level

Relation Between Flying Training Performance and Personality Factor Scores

For a subsample of 144 respondents with check flight grades, composite performance measures were generated separately for the T-37 and T-38 phase of training. The composite check flight grade was based on the mean of the three check flight grades obtained within each phase. The correlations between the personality factor scores and the mean check flight grades are shown in Table 5. These data suggest that self-confidence, although previously shown to be associated with pass/fail, was not significantly related to performance either in the T-37 or T-38 phases. In contrast,

Table 5 Correlations Between Personality Factor Scores and Flying Training Performance ($\underline{N}$ = 144)

| | Performance Measure | | |
|---|---|---|---|
| Personality Factor Score | T-37 | T-38 | Combined |
| Hostility | -.05 | -.11 | -.10 |
| Self-confidence | .11 | .08 | .12 |
| Values Flexibility | .09 | .02 | .06 |
| Depression | .00 | $-.26^*$ | $-.18^*$ |
| Activity Level | $-.17^*$ | -.12 | $-.18^*$ |

Note: T-37 represents first phase of UPT flying training; T-38 represents the second phase of UPT flying training. Combined represents the average of both phases of training.

* $p < .05$ that variation from zero correlation is due to chance.

the depression measure, although not associated with graduation, was negatively and significantly correlated with performance in the T-38 phase ($r$ = -.26, p < .05). Activity level was also somewhat negatively associated with check flight performance, particularly in the T-37 phase of training ($r$ = -.17, $p$ < .05).

## DISCUSSION

The main finding to result from the present study was that the use of global personality scores could be used to enhance the predictive validity of the current selection system. However, a second finding was that choice of criteria may have an important influence on deciding upon the optimal set of predictor measures. That is, measures related to training attrition concerns (pass/fail) may not be predictive of training performance, and measures of performance at different phases of training may be differentially associated with selection measures.

Some of the specific results of the present research should be interpreted cautiously. The non-graduate groups consisted of a fairly small number of respondents (from ten to 34); moreover, reasons for elimination may not represent mutually exclusive categories. That is, a student performing poorly might initiate his or her own elimination (SIE) rather than be terminated for poor performance (FTD), in order to avoid potential embarrassment. Similarly, a poor performer might discover medical problems or develop manifestations of apprehension (MOA) rather than be eliminated as an FTD. Thus, considering both the sample sizes and the nature of the criteria, the relationships found between the personality measures and the elimination groups should best be regarded as suggestive rather than conclusive.

Current research efforts, designed to support a new direction in Air Force pilot training philosophy, involve personality measures associated with effective performance in crew operations. As currently proposed, candidates for U.S. Air Force Specialized Undergraduate Pilot Training (SUPT), to begin in 1991, would be assigned to either a fighter/bomber track or a transport/tanker track before entry into any phase of USAF pilot training, rather than after 42 weeks of training, as is the current practice. Consequently, the Air Force Air Training Command, the organization responsible for pilot training, is interested in research to develop predictors of which pilot candidates will best be suited to the proposed training tracks. Currently, a battery that includes personality traits associated with effective crew coordination, and also a measure of self-confidence, are being administered in a group of pilot candidates designated to attend pilot training in the near future.

To conclude, the present research provided evidence for the utility of personality measures in the current Air Force pilot well vary depending on the criteria selected for evaluation (training graudation versus training performance). Finally, current efforts are underway to replicate these findings and examine the use of such measures for a proposed new pilot selection and classification system.

## REFERENCES

[1] North, R. A., & Griffin, G. R. Naval Aerospace Medical Research Laboratory. Aviation selection 1919-1977, 1977, Special Report No. 77-2.

[2] Imhoff, D.L. & Levine, J.M. Manpower and Personnel Division, Air Force Human Resources Laboratory. Perceptual-motor and cognitive performance task battery for pilot selection, 1981, AFHRL-TR-80-27.

[3] Carretta, T. R. Manpower and Personnel Division, Air Force Human Resources Laboratory. Basic Attributes Tests (BAT) System: A Preliminary evaluation, 1987, AFHRL-TR-87-20.

[4] Dolgin, D. & Gibb, G. D. Personality Assessment in Aviator Selection. In Jensen, R. (Ed.), Aviation Psychology: 3rd Edition. London: Gower Publishing Group, 1988, 285-319.

[5] Sells, S. B. Development of a personality test battery for psychiatric screening of flying personnel. Journal of Aviation Medicine, 26, 1955, 35-45.

[6] Jessup, G., & Jessup, H. Validity of the Eysenck Personality Inventory in pilot selection. Occupational Psychology, 45, 1971, 111-123.

[7] Novello, J. R., & Youssef, Z. I. Psycho-social studies in general aviation : I. Personality profile of male pilots. Aerospace Medicine, 45, 1974, 185-188.

[8] Novello, J. R., & Youssef, Z. I. Psycho-social studies in general aviation: II. Personality profile of female pilots. Aerospace Medicine, 45, 1974, 660-663.

[9] Ashman, A., & Telfer, R. Personality profiles of pilots. Aviation, Space, and Environmental Medicine, 54, 1983, 940-943.

[10] Bartram, D., & Dale, H. C. A. The Eysenck Personality Inventory as a selection test for military pilots. Journal of Occupational Psychology, 55, 1982, 287-296.

[11] Fry, G. E., & Reinhardt, R. F. Personality characteristics of jet pilots as measured by the Edwards Personal Preference Schedule. Aerospace Medicine, 40, 1969, 484-486.

[12] Sanders, M. G., & Hoffman, M. A. Personality aspects of involvement in pilot-error accidents. Aviation, Space, and Environmental Medicine, 46, 1975, 186-190.

[13] Sanders, M. G., Hoffman, M. A., & Neese, T. A. Cross-validation study of the personality aspects of involvement in pilot-error accidents. Aviation, Space, and Environmental Medicine, 47, 1976, 177-179.

[14] Wichman, H. & Ball, J. Locus of control, self-serving biases, and attitudes towards safety in general aviation pilots. Aviation, Space, and Environmental Medicine, 54, 1983, 507-510.

[15] Jones, A. Navy Scientific Advisory Group. A survey of military pilot selection procedures in ten countries, 1983, SP(N) Report R56.

[16] Skinner, M.J. & Ree, M.J. Manpower and Personnel Division, Air Force Human Resources Laboratory. Air Force Officer Qualifying Tests (AFOQT): Item and factor analysis of Form O, 1987, AFHRL-TR-86-68.

[17] Dahlstrom, W.G., Welsh, G.S. & Dahlstrom, L.E. An MMPI Handbook Vols. I & II. Minneapolis: University of Minnesota Press, 1972.

[18] Bucky, S. F., & Spielberger, C. D. State and trait anxiety in voluntary withdrawal of student naval aviators from flight training. Psychology Reports, 33, 1973, 351-354.

[19] Knapp, R.R. Handbook for the Personal Orientation Inventory. San Diego: Edits, 1975.

[20] Mauger, P.A. & Adkinson, M.A. Interpersonal Behavior Survey (IBS) Manual. Los Angeles: Western Psychological Services, 1980.

[21] Jenkins, C.D., Rosenman, R.H. & Friedman, M. Development of an objective psychological test for the determination of the coronary-prone behavior pattern in employed men. Journal of Chronic Diseases, 20, 1967, 371-377.

[22] Siem, F. M. Manpower and Personnel Division, Air Force Human Resources Laboratory. Characteristics associated with success in USAF pilot training, submitted.

[23] Goldstein, M. Maximal correlation between normal and dichotomous variables. Journal of Educational Statistics, 11, 1986, 259-261.

## DISCUSSION

BILLINGS: If I was able to read the slides correctly, it would appear that the highest correlations you obtained accounted for somewhere in the neighbourhood of 3-4% of the variance on the measures. I wonder if that may be one of the reasons why we stopped taking these sorts of measures some considerable number of years ago. Granted they are statistically significant, but can we possibly ascribe any sort of operational significance to correlation co-efficients in those area?

SIEM: I think it is misleading to look at those point by serial correlations and try to interprete them as very significant because, given the ratio of graduates and non-graduates in the sample, the maximum correlation possible would be 0.7. It is hard to interpret a correlation co-efficient that is not on a zero to one scale. That is one factor. I think the fact that there was incremental validity in the analyses, described in the paper, does seem to improve selection. Obviously we do not use these personality measures by themselves but we build them into the selection system. I think the incremental validity does provide some justification for at least giving these factors some weight in a total selection system.

SIOMOPOULOS: I would like clarification about the component called 'flexibility'. We know that the ability to think abstractly is important in order to have flexibility. So were you checking flexibility or were you checking the ability of the pilot to think abstractly? In my experience I think that pilots need to be able to think about abstract operations; if they are only concrete thinkers they cannot be really good pilots.

SIEM: I had no idea that this was an empirical relationship. Frankly, I find it most intriguing and I would definitely like to follow it up. It is exciting because it is a kind of bridge between the cognitive and the personality areas of psychology which makes it a finding worth pursuing.

# Leader Personality and Crew Effectiveness: Factors Influencing Performance in Full-Mission Air Transport Simulation

Thomas R. Chidester
H. Clayton Foushee

NASA-Ames Research Center
Mail Stop 239-15
Moffett Field, CA 94035
U. S. A.

## SUMMARY

A full-mission simulation research study was completed to assess the potential for selection along dimensions of personality. Using a selection algorithm described by Chidester (1987), captains were classified as fitting one of three profiles using a battery of personality assessment scales, and the performances of 23 crews led by captains fitting each profile were contrasted over a one and one-half day simulated trip. Crews led by captains fitting a Positive Instrumental-Expressive profile (high achievement motivation and interpersonal skill) were consistently effective and made fewer errors. Crews led by captains fitting a Negative Communion profile (below average achievement motivation, negative expressive style, such as complaining) were consistently less effective and made more errors. Crews led by captains fitting a Negative Instrumental profile (high levels of Competitiveness, Verbal Aggressiveness, and Impatience and Irritability) were less effective on the first day but equal to the best on the second day. These results underscore the importance of stable personality variables as predictors of team coordination and performance.

## INTRODUCTION

Crew effectiveness is a joint product of the piloting skills, the attitudes, and the personality characteristics of team members. As obvious as this point might seem, traditional approaches to the optimization of crew performance have emphasized the skills dimensions almost exclusively. More recently, many crew training programs have been expanded to include the influence of interpersonal characteristics associated with crew coordination in Cockpit Resource Management (CRM) programs (see Orlady & Foushee, 1987). While it is certainly encouraging that efforts are now underway to improve both the technical and interpersonal skills dimensions, Helmreich (1984) has argued that individual differences such as stable personality characteristics are unlikely to be impacted by training approaches.

Unfortunately, the search for stable personality predictors of performance has been plagued by a historic failure to validate links between personality and performance. The reasons for this failure are complex and beyond the scope of this paper. However, one reason is that most studies have focused on performance parameters during very limited periods of time, such as after the completion of a training program (Melton, 1947). Helmreich, Sawin, and Carsrud (1986) argued that performance during and shortly after training is much less sensitive to personality effects than performance after the effects of training have begun to subside. These researchers tracked a group of airline employees during training and for their first six months on the job. Personality predictors did not significantly correlate with performance evaluations made during and immediately after training, but after six months, the personality predictors became significantly linked to performance. Helmreich, et al. (1986) interpreted these data as evidence of a "honeymoon effect." During training and the first days on the job, most individuals are motivated to do as well as possible. Over time, the job settles into a routine and personality characteristics such as "intrinsic motivation" seem to become more important predictors of performance.

Our strategy for evaluating the potential of crew selection on the basis of personality characteristics was to identify dimensions of personality theoretically linked to performance in group settings and to conduct a high-fidelity validation study to determine whether these theoretical relationships translate into real-world performance. Even though our understanding of group phenomena is not what it should be, there are a number of possible criteria that have been suggested by previous research. First, high levels of individual technical skill, proficiency, and the motivation to work hard are the foundation upon which effective crew coordination is built. Second, past research (e.g. Foushee & Manos, 1981; Foushee, Lauber, Baetge, & Acomb, 1986; Kanki, Lozito, & Foushee, 1988) has demonstrated that effective crews are characterized by frequent, direct, open, and concise communication. Third, effective leadership is a joint function of: a) effective task delegation and definition; b) effective cross-checking and feedback; and c) creating an atmosphere where subordinates feel free to offer suggestions and counter-proposals to leader-prescribed courses of action. Our selection task was to seek measures that optimized these features.

A great deal of emphasis has been placed upon the first dimension, individual technical skills and the motivation to achieve. For selection purposes, we chose to focus on the motivational component of individual performance or overall level of "instrumentality", which we operationally define as a person's goal orientation and independence. We also chose to emphasize the overall level of "Achievement Striving" as an additional measure of an individual's dispositional orientation toward task performance situations. The second dimension is oriented toward communication and interpersonal exchange, and we focused upon dimensions that would theoretically facilitate communication between group members. This dimension is commonly defined as "Expressivity," or interpersonal warmth and sensitivity. Conversely, communication would be expected to be inhibited in groups composed of individuals with styles characterized by both negative expressive (frequent complaining, etc.) and negative instrumental traits such as Verbal Aggressiveness, Competitiveness, or Impatience and Irritability. In summary, we theorized that effective leaders are more often characterized by relatively high levels of both Positive Instrumentality and Positive Expressivity (high levels of both concern for people and concern for performance), and that this type of leadership style would facilitate crew performance. Lower levels of Positive Expressivity and higher levels of negative expressive and negative instrumental traits were expected to lead to less effective crew communication, coordination, and performance overall.

We chose these dimensions for a number of reasons. A number of personality theorists and researchers have focused on dimensions reflecting Instrumentality and Expressivity as one, if not the most, central set of personality characteristics. Influential theorists (e.g. Spence & Helmreich, 1978; Bakan, 1971; Fiedler, 1967; Angyal, 1965; and Halpin, 1954) have all in one way or another identified these characteristics as core components of human personality with strong behavioral relationships. Moreover, many popular management theories have espoused concern for people balanced with concern for performance as the key to leader success. Blake and Mouton's (1978) "managerial grid" is perhaps the most widely applied example of these notions, and it has been incorporated into many training programs, including a number in aviation.

Also interesting is the fact that many managerial training programs emphasize these dimensions and seek to train individuals to adopt these characteristics for increased managerial effectiveness. Unfortunately, there is little if any evidence to indicate that short-term training programs produce any substantive behavioral change in these areas over the long term. It may be difficult to teach individuals to adopt behaviors such as these because they may be linked to stable personality traits that, by definition, are relatively enduring and shaped over long periods of time. To the extent that these desirable managerial characteristics are related to stable traits, training will not be an effective means of address. The only effective approach would be a balance of both training and selection techniques--selecting individuals with appropriate constellations of traits and training in effective crew coordination and group problem solving techniques.

Since there are numerous dimensions at least theoretically relevant to crew effectiveness, the selection algorithm had to allow for appropriate combinations of trait dimensions. There are at least two ways to approach this problem. The first is based on theory and calls for a simple combination of dimensions that are either theoretically or empirically associated with superior performance by selecting only candidates at the upper extreme of all dimensions. However, this strategy usually proves somewhat impractical because when numerous dimensions are involved, most of the population is eliminated. An alternative strategy is to consider not only relationships between each predictor and performance, but also the interrelationships among predictors. This strategy takes into account the frequencies or "clusters" of individual trait dimensions that tend to normally co-occur within individual people. Some combinations may be rare and thus difficult to select in an average population. For example, it would be difficult to imagine traits associated with Verbal Aggressiveness occurring alongside of those related to shyness in a normal individual. We chose this "cluster" approach as our selection strategy.

The selection algorithm utilized for this study was developed along with Robert Helmreich and colleagues at the University of Texas (see Chidester, 1987, for details on specific selection instruments). Cluster analyses were accomplished on two samples of pilots, who had been administered the various personality batteries, as a means of identifying groups of individuals with similar standing on the multiple dimensions. Cluster analysis is a statistical technique which combines subjects into groups or clusters based upon each subject's similarity to other subjects along any specified set of dimensions (Norusis, 1988). Sample one consisted of civilian airline pilots, and sample two was made up of military pilots. Three distinct clusters were found in both samples, one with high levels of positive traits and two others with different constellations of negative traits. Pilots in the positive cluster were characterized by high levels of Instrumentality, Expressivity, and Achievement Striving (Work and Mastery) and were designated the Positive Instrumental-Expressive or "IE+" cluster. One of the negative clusters was characterized by high levels of Negative Expressivity and low levels of Instrumentality and Achievement Striving. This cluster is characterized by traits associated with tendencies to express oneself in a negative fashion (e.g. complaining) and lower than average goal orientation. It was labeled the Negative Expressive or "Ec-" cluster. The second negative cluster was characterized by higher than average levels of Verbal Aggressiveness, Negative Instrumentality, and Competitiveness. This cluster comprises a more "authoritarian" orientation and may well be associated with elements of a profile popularly known as "the right stuff." It was labeled the Negative Instrumental or "I-" cluster.

There is also some existing evidence that these clusters are related to crew performance dimensions. Chidester (1987) found that pilots responded differentially to training in crew coordination as a function of these profiles. IE+ pilots benefitted the most from training as assessed by cockpit management attitude change. The current study was designed to: 1) evaluate whether the personality characteristics of individual crewmembers significantly impact the crew performance process and; 2) evaluate the experimental selection algorithm as a possible countermeasure for the prevention of crew coordination problems. A two-day full-mission simulation study was designed, in which crews flew 5 flight segments under varying conditions of workload and in which crews were composed differently according to the selection criteria previously described. Specifically, the captain of each crew was randomly selected from one of the three clusters: IE+, I-, or Ec-. Data were collected from a number of sources, including self-reports, expert observation, video-coding of errors, aircraft handling parameters, and audio recording of flightcrew and air traffic communications.

## METHOD

### *Design and Procedures*

Twenty-three, three-person crews (69 pilots) completed a one and one-half day full-mission simulation of airline operations in the Ames Man-Vehicle Systems Research Facility (MVSRF) B-727 simulator. All crews were employed by the same major U.S. air carrier, all crewmembers were currently operating the B-727 exclusively in passenger operations, and all crewmembers were at the time qualified in the B-727 crew position (e.g. captain, first officer, second officer or flight engineer) they occupied in the simulation. Three different types of crews were composed. The crew types contrasted were based upon cluster membership as described by Chidester (1987).

Crew types represent selection for leadership; that is, only the captain's personality characteristics were considered when crews were composed. The first crew type was composed of a randomly assigned first officer and flight engineer flying with a captain from the IE+ cluster. We expected these captains to be good leaders and their leadership to translate into effective crew performance. The second type was composed of a randomly assigned first officer and flight engineer flying with a captain from the I- cluster. The third type was composed of a randomly assigned first officer and flight engineer flying with a captain from the Ec- cluster. These two negative leader crews were expected to be less effective at crew coordination in the high workload flight segments.

Crews flew 5 flight segments. Each segment was planned and flown as closely as possible to real operations. Crews were provided with all of their normal flight documentation, completed all normal flight and cockpit preparations, and communicated with all ground support personnel normally available to them (dispatch, ground crew, air traffic control, ATIS, and maintenance). Flight routings corresponded to typical clearances along routes in central and southern California. Segments 1-3 were flown on the first day, segments 4 and 5 occurred on the second day. Routine levels of workload were designed into segments 1,2, and 4, but segments 3 and 5 were far more demanding than normal and involved continuing abnormal conditions that could not be resolved completely in flight. In short, the task involved a complex process of diagnosis, assessment of options, and coordination of procedures required to land an aircraft with mechanical problems.

Past full-mission simulation research (Ruffell-Smith, 1979; Lauber & Foushee, 1981; Foushee, et. al., 1986) has shown that successful crew performance simulation scenarios have at least five essential elements. First, they are designed to be completely representative of the actual operational environment, and all details are faithfully represented. Second, they are complicated enough to require the coordinated action of all crewmembers for successful completion, but not to the extent that they induce complete crew failure such as a "crash." Third, problems presented to crews have ongoing consequences which must be dealt with in flight, but cannot be fixed in flight. Fourth, the problems involved are very ambiguous, and there is usually no simple corrective "by the book" solution. And fifth, the original problem is usually compounded by other events such as weather-induced complications (e.g. landing on a rain-slick runway with partial brake failure). It is also interesting to note that these characteristics have been seen in past incidents and accidents.

In the process of scenario design, outlines of potential events were developed by the principal investigators using accident case studies and incident reports. These were reviewed by individuals with checking and training experience in the particular aircraft type being simulated and by simulator operational personnel from the MVSRF. Typical environmental conditions for the proposed area of flight (November-February weather patterns for coastal and central California) were considered in great detail, so that weather patterns and scenario events would be realistic to the experimental flight crews. Aircraft documentation and airline dispatch procedures were assembled for each flight segment in cooperation with airline management and the local pilot labor union executive committee.

Following this development process, selected scenario outlines were programmed into the simulator computer and eight pre-test runs were conducted using qualified flight crews to refine procedures, train facility personnel and the experiment staff, and test scenario events. These pre-test crews were carefully debriefed to assess the realism of the scenarios and procedures used by the experiment staff. This feedback allowed continual refinement until the scenarios were finalized.

The scenarios involved operations within California in the U.S. and were adapted from those of an earlier NASA investigation. Airfields available to the crew included San Francisco (SFO), Stockton (SCK), Sacramento (SMF), and Los Angeles (LAX). After completing 2 familiarization segments, crews flew a total of 5 legs between pairs of these cities, 3 on the first day and 2 on the second day. Test scenarios (legs into which abnormalities were inserted and full simulator data collected) occurred on the third leg of the first day and the second leg of the second day. Crews were led to believe, however, that they would fly 6 legs. The last leg (SMF-SFO) was not intended to occur, except as a part of preflight planning, due to scenario events. This deception was intended to counteract suspicions that might be associated with the last segment of the study, particularly since the last segment of day one involved an abnormality.

On the first day, all legs except the third were relatively routine. However, one irregular item was included in the dispatch paperwork for the full-mission segments. The deferred item list and accompanying minimum equipment list (MEL), indicated that the plane was to be dispatched with the #3 generator inoperative. This was a legal procedure, but given the weather conditions (night, fog, low ceilings), prudent crews should have considered delaying departure until either the weather improved or the generator was repaired. If the crew elected to request the repair, the generator was "repaired", but the dispatcher warned the crew that they might expect it to malfunction again requiring it to be reset. If the crew did not request repair, the generator was "repaired" later when the crew reached the maintenance base at SFO. In either case each time the #3 engine was started following the "repair", the generator field light illuminated, and at least one field reset was required to bring the generator on line. This manipulation was intended to complicate the decision-making process on a later high workload segment.

On segment #3 (day one test segment), crews flew from Sacramento to Los Angeles. Following a normal take-off and climb, a combination of system failures were activated automatically. First, the vertical stabilizer trim system began running uncommanded and jammed in a nose-down condition at a predetermined point. Second, and shortly thereafter, when the aircraft crossed a specified navigational point, the #2 engine low oil pressure warning light illuminated and the indicated pressure fell to the cautionary range. This combination represented a relatively high-workload situation, but was compounded by neither weather, traffic, nor ATC problems. No diversion from flight plan was necessary, except for actions required to stabilize the aircraft and land safely.

The scenario presented the crew with two independent failures which could have impacted flight safety. The jammed stabilizer trim system was a serious control problem, disabling the autopilot, and requiring (due to the descent configuration) constant nose-up control inputs and considerable back-pressure from the flying pilot. This, of course, made the approach and landing much more difficult and was physically fatiguing (given the need to hold constant, firm back-pressure). The procedure for dealing with the jammed stabilizer trim, once it is identified was to limit flap setting to 15 degrees for landing, increase approach speed by 15 knots, and establish the landing configuration as early as possible so as to get a feel for the control forces necessary for landing. This procedure is difficult but not unreasonable since it is required for all type ratings in the B-727 aircraft. The low oil pressure light and corresponding cautionary gauge reading were also covered by a checklist procedure, but the outcome of the checklist left the crew a choice. At the captain's discretion, the crew could either shut the engine down or reduce thrust on the engine to idle. Shutting the engine down required the completion of more checklists (greatly increasing workload and time requirements), may have required the dumping of fuel (again increasing workload), and removed one generator from operation. Recall that the crew had already encountered minor problems with the #3 generator, so the aircraft could have easily been down to one should conditions have deteriorated further. The most prudent course, then, was to keep the engine operating as a reserve while continually monitoring its operating parameters. Once these decisions were made, the crew had to land the aircraft in an abnormal configuration.

On the second day test segment, the weather continued to be characterized by fog, overcast, and minimal visibilities in the central California valley. The first leg (SCK-LAX) went without programmed incident except for fog in the Los Angeles area (ceiling 400 ft., visibility 1 mile). By the time of the second leg (LAX-SMF), weather had deteriorated further in the central valley, resulting in poor visibility in the Sacramento area (ceiling 300 ft., visibility 1/2 mile; the legal minimums for Category 1 approaches). Weather in Los Angeles remained foggy but above legal minimums. Following a normal departure and climb out of the Los Angeles area and normal cruise, the crew received clearance for the Wraps Four Arrival to Sacramento. As the crew entered the Sacramento terminal area, runway visual range (RVR) was reported by the approach controller as 2000 ft., again just above the Category 1 approach minimums. Prudent crews should, at this point, have considered and briefed for a possible Category 2 approach and landing. The minimums for this approach were RVR 1200 with a decision height of 126 ft. However, after the aircraft crossed the outer marker, RVR was reported by the tower as less than 1000 ft., which was below the minimums for the initiation of an approach. However, since they were inside the marker, it was legal to continue until the published decision height and to land if the captain had a sufficient view of the runway environment. If the crew attempted this, the runway was not visible at the decision height, requiring the approach to be aborted. During the missed approach, as the crew retracted the flaps, a hydraulic system A failure occurred, caused by a leak that depleted all of the hydraulic fluid. At this point, it was immediately apparent that a diversion to an alternate airport would be necessary. Weather conditions at various nearby alternates were poor (all ceilings less than 800 ft. and visibilities less than 1 mile). Reno, San Francisco, Oakland, and San Jose (SJC) were the best alternates. Reno was unacceptable because the prevailing wind direction made its shortest runway the active runway, and the use of this runway was prohibited by company policy. SFO had the best weather of the remaining alternates, with clearing conditions, a 2000 ft. ceiling, and 2 miles visibility with light rain. SFO also provided a very long runway at just under 12,000 ft. However, SJC was listed as the flight's legal alternate, because SFO weather was below alternate minimums at the time of dispatch.

This scenario confronted the crew with a number of hazards and limitations. First, the hydraulic failure disabled a number of aircraft systems. The landing gear had to be extended by hand crank, and once extended could not be retracted. Flaps also had to be extended by alternate means, and this system does not allow leading edge flaps to be retracted once extended. In addition, the trailing edge flaps were not protected against asymmetric extension using the alternate system. Alternate extension required more time than extension by normal means, and was limited to 15 degrees in case of a missed approach (flap retraction from the normal 30 degree setting by the alternate system following a missed approach would be too slow to reduce drag sufficiently to allow the aircraft to climb to obstacle clearance altitude). The hydraulic failure also disabled nose wheel steering, ground spoilers, and outboard flight spoilers. All of these limitations caused a combination of higher than normal approach speeds and reduced stopping ability. Finally, the crew had to select an airport (SFO is suggested by the circumstances) and execute an approach and landing under adverse circumstances. When the crew extended the flaps on approach to SFO, they received an outboard trailing-edge flap asymmetry indication resulting from the lack of protection discussed above. This condition changed the handling characteristics of the aircraft and required that crews discuss and estimate a landing speed, because none is given in the flight manuals for the condition of split inboard/outboard flaps combined with a hydraulic system failure.

*Measurement*

Prior to scheduling for the simulation, candidate subject pilots completed the personality selection battery. The battery was scored so that the cluster profile could be determined for each crewmember. Crew performance data was collected from three sources: expert observations, video-coding of crew errors, and computer recording of aircraft handling parameters. A recently-retired, highly experienced airline captain served as the expert observer and was present in the simulator cab with every flight crew. He was blind to the experimental condition, and evaluated crew performance following every flight segment and individual performance during specific phases of the high-workload segments.

Error analyses were undertaken using two independent sources of data to assure the reliability of performance assessment. First, during test runs, the expert observer kept a record of all errors he observed. The second source of error data required a complete review of the videotape records. Using these records, two condition-blind observers reviewed each flight for operational errors. When an error was recognized by one or both observers, the tape was stopped and the segment containing an alleged error reviewed. After this process, both observers agreed that the error occurred or it was not counted in the analysis. All errors identified by the videotape observers were then presented to the expert observer, who had the option of eliminating an error on the basis of his notes taken during the simulations or his operational experience. This was a very conservative error tabulation process and assured that every error data point was reviewed at least twice.

Since some performance errors were more operationally significant than others, errors were categorized according to level of severity. This process was accomplished by the expert observer and by both of the observers involved in the videotape error analysis. A three-level classification was utilized. Type 1 errors were defined as minor, with a low probability of serious flight safety consequences. Type 2 errors were defined as of moderate severity, with a stronger potential for flight safety consequences. Type 3 errors were defined as major, operationally significant errors having a direct negative impact upon flight safety.

Two additional types of data were collected but have not yet been processed. Aircraft handling data will allow the assessment of deviations from prescribed paths during specific flight phases, particularly instrument approaches during the high workload segments. Finally, transcripts of all communications within crews and between crews and air traffic control will be produced. Each communication will be coded by speaker, to whom it was directed, and content. These communications data will be used to examine the crew coordination process in great detail. It is expected that these data will be the most sensitive to the effect of the selection criteria and that the communications process is the mediator of links between individual differences and performance.

## RESULTS

*Observer Ratings*

Analyses of the observer's ratings of crew performance during the full-mission segments revealed a significant interaction between leader personality and flight segment ($F(8,80)=2.80$, $p<.01$). Simple-effects tests were conducted for between-crews and within-crews cell means. Means for each group of captains during each segment are presented in Table 1.

**Table 1 - Observer Ratings of Crew Performance**

| Captain Personality | Flight Segment | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| IE + | 3.85 | 3.88 | 3.95* | 4.37 | 4.22 |
| I - # | 3.40 | 3.55 | 2.97 | 3.79 | 3.98 |
| Ec - | 3.23 | 3.59 | 2.90 | 3.12 | 2.90* |

* Indicates significant between-crew differences

# The following within-crews comparisons are significant among I- led crews: 5 vs. 1, 2, & 3; 3 vs. 4.

Examination of these means reveals that crews led by IE+ captains were rated as consistently effective, and these ratings were higher than the other crew types for the segments overall (though not every comparison for every segment is statistically significant). Crews led by Ec- captains were rated as consistently less effective over all segments than those led by IE+ captains (though not all comparisons were statistically significant). Crews led by I- captains received ratings that varied considerably across segments. For segments 1, 2, and 3, I- led crews were similar to Ec- crews; they were rated as less effective than IE+ crews. However, on segment 5, I- led crews were rated as performing as well as IE+ led crews, and significantly more effectively than Ec- led crews. A number of comparisons were significant over time for I- led crews. Performance on segment 5 was rated as more effective than on segments 1, 2, and 3. Performance on segment 3 was rated as less effective than segment 4.

*Error Analyses*

Error analyses had been completed for 19 of the 23 crews at the time of this report. A total of 833 errors were identified by the video observers across these 19 crews. Of that number, 74 (8.9%) were eliminated by the expert observer as logical choices made by the crew in response to available information. Of those errors that were eliminated, 53 (72%) were initially rated as Type 1 errors by the video observers; the remainder were initially rated as Type 2 errors. None of the eliminated errors was initially rated as a Type 3.

A consensus severity classification was reached by the video and expert observers for the remaining 759 errors. Table 2 shows the relative frequency of each type of error.

**Table 2 - Frequency of Errors by Severity Classification**

| | Type 1 | Type 2 | Type 3 |
|---|---|---|---|
| Count | 337 | 236 | 186 |
| Percent | 44 | 31 | 24 |

In order to emphasize the distinction between errors of no consequence and significant operational errors, error counts for each crew were submitted to 2 separate analyses. The first analysis included Type 1 error counts for each day; the second included Type 2 and Type 3 error counts, with severity level entered as a design factor. To control for the impact of a significant flight time difference between segment 3 and segment 5 ($F(1,16) = 179.3$, $p < .001$; means = 66 and 102 minutes, respectively), and the correspondingly increased opportunity to make errors on segment 5, flight time was entered as a covariate for each day for each crew. Flight time was unrelated to the experimental variables.

Analyses of Type 1 errors revealed no significant main effects or interactions. Type 1 errors appear to be randomly distributed across crews. Analyses of Type 2 and Type 3 errors revealed two main effects. First, crews tended to make more Type 2 (moderate) than Type 3 (major) errors ($F(1,15) = 4.98$, $p < .05$; means = 6.4 and 5.0, respectively). Second, Ec- led crews tended to make more errors than IE+ or I- led crews ($F(2,15) = 4.44$, $p < .05$; means= 13.7, 9.4, and 8.9, respectively; post hoc comparisons via Tukey's HSD).

While these error analyses were consistent with ratings by the expert observer in discriminating performance of Ec- led crews from IE+ or I- led crews, they did not reveal a change in I- led crew performance over the course of the simulation. That is, the Captain Personality by Flight Segment interaction, significant for expert observer ratings, was not for crew errors. One possible explanation for this discrepancy is that in the process of systematically rating crew performance, observers may take more than errors into account. This might be expected, since the observer has access not only to errors, but also to the crew communications process which is a significant predictor of crew performance (Foushee & Manos, 1981). Completion of the process analyses examining patterns of communication may shed light on this inconsistency.

*Results Summary*

These analyses suggest divergent patterns of performance as a function of leader personality type. IE+ led crews show consistently effective performance. Ec- led crews show consistently less effective performance. I- crews show a familiarity effect, or improvement over time--relatively less effective performance on the first day and relatively effective performance on the second day. In addition, these results are not of trivial magnitude. The effect size (Rosenthal, 1978) for rated performance of an IE+ versus an Ec- led crew, for example, is .97 (standard deviations), averaged over high and low workload flight segments. For crew errors, the effect size statistics are -.80 (segment 3) and -1.05 (segment 5) for Type 2 errors, and -.46 and -.70 for Type 3 errors. Cohen (1977) has shown that small, medium, and large effect sizes in psychological research correspond to approximately .2, .5, and .8 standard deviation units, respectively. This suggests a robust effect, which is even more credible given the generalizability of full-mission simulation performance to the operational environment.

## DISCUSSION

These results present a consistent picture of the impact of leader personality on crew performance. Analyses of the crew effectiveness ratings by an expert observer and of the crew error frequencies revealed three patterns of crew performance as a function of leader personality. Comparing our results to our hypotheses, we found overall support with some exceptions. We expected consistently effective performance from crews led by IE+ captains. We expected less effective performance from crews led by I- or Ec- captains. We did not expect I- led crews to perform comparably with IE+ crews in general, or that I- led crews would apparently improve over time. From previous research (Foushee, et al, 1986), we expected that increasing crew familiarity would result in better crew performance in the later flight segments, regardless of crew type. Instead, we found that familiarity apparently facilitated performance only in I- led crews. Familiarity did not appear to enhance performance in the IE+ group, however this may have been due to a "ceiling effect" of sorts. These crews were obviously performing at a relatively high level from the beginning. Moreover, familiarity did not produce improved performance in the Ec- crews. Their performance was consistently less effective over all segments. It is possible that the improvement seen in I- crews was sufficiently strong to produce the overall familiarity effect seen in the Foushee et al. (1986) study where leadership style variables were not experimentally manipulated.

The findings are intriguing because they support logical hypotheses about crew effectiveness and are consistent with recent results found for individual pilot performance. Foushee and Helmreich (1988) have hypothesized that relatively high levels of both instrumental and expressive traits might facilitate crew performance. IE+ individuals have elevated levels of both of these traits, and as noted earlier, Chidester (1987) found that IE+ pilots appeared to benefit most from CRM training in a sample of military pilots. This group showed the most positive and enduring attitude change up to 6 months after participating in the training program. The results from this study provide further support for the notion that both Instrumentality and Expressivity are important predictors of performance in team environments.

*Crew Familiarity*

The familiarity effect among I- led crews raises a number of important questions. Why was the pattern of lower performance seen in the observer ratings not reflected in the errors committed by I- led crews on the first day? The expert observer saw these crews as relatively ineffective, giving them the lowest average ratings for segment 3. However, these crews made no more errors than IE+ led crews. One reason may be that process observers are, by definition, integrating more than errors into their observations in any group task situation. In previous studies (Ruffell-Smith, 1979; Foushee & Manos, 1981; Foushee, et al., 1986; Kanki, Lozito, & Foushee, 1988), patterns of flight crew communications are significant predictors of crew performance. As a result, we have incorporated communications dimensions into our observer rating scales. These scales ask the observer to use his experience and professional judgment to evaluate how crews make decisions, handle inter- and intra-crew communications, prioritize problems and distractions, and distribute workload. In short, these ratings seek to evaluate the *process* by which crewmembers coordinate their activities. Thus, it should not be expected that ratings reflecting process dimensions be perfectly correlated with performance outcomes. They are related, but they are not the same thing. Problems of crew coordination do not always produce observable errors, but are important in their own right because these types of group process problems certainly raise the probability that errors will be committed or not corrected quickly. We would argue that our observer ratings reflect the fact that there were significant process problems within I- crews.

This argument is also consistent with the conceptual framework proposed by McGrath (1964, 1984). In this model, the link between input variables (such as personality profiles) and group outcomes (such as errors of communication or action) is mediated by the process of group activities (such as patterns of communication). As a result, it is possible to identify links between input and process or process and outcome variables that diverge somewhat from input-outcome relationships. The observer ratings may be viewed as integrating both process and outcome information.

The idea that familiarity may have affected the process of crew interaction in I- crews raises another interesting question. What behaviors were changing over the course of the simulation? *There are at least two* possibilities. One is that the leader may be altering his behavior following the high workload segment at the end of the first day, while the other suggests that the crew may be adapting their behavior to the captain's expectations. We suspect that the latter interpretation is more plausible. In air transport operations, it is not unusual for a crew to be composed of individuals who have never met prior to the beginning of a trip. Accordingly, an adjustment period is likely during the first few flight segments, and it is probable that subordinate crewmembers often attempt to tailor their behaviors to the captain's expectations. Since the captain's role is more central to cockpit organization than other positions, we suspect that a captain is less likely to change his behavior to adapt to the crew (although this may occur to a lesser degree as well).

Foushee, et al. (1986) did demonstrate significant process differences between crews that had recently flown together versus those that had not. It seems reasonable to speculate that initially, all subordinate crew members are tentative in their behavior because they are awaiting signals from the leader about how he or she expects the cockpit to operate. In general, IE+ leaders would be expected to very quickly create an atmosphere where open communication is encouraged. Theoretically, I- leaders would not be as likely to do so and might by nature tend to discourage questioning by subordinates. After the initial adjustment process, subordinates in I- crews may have been able to work more effectively because they knew what to expect. However, tangible evidence of this change in the I- crews will have to await the process analyses in this study.

Another important question is related to the generality of this familiarity effect and its application over time. The current study and the Foushee et al. (1986) study only compared crews who had worked together for two or three days, and familiarity seemed to provide a performance benefit in a substantial number of these cases. However, we know little about team performance over longer durations, and it is quite possible that increasing familiarity could ultimately result in worse performance. In the current research, we have suggested that a number of attributes of I- leaders might be viewed as aversive under certain circumstances. I- leaders are characterized by high levels of Impatience and Irritability, Competitiveness, and Verbal Aggressiveness combined with low levels of Expressivity. So far we have shown that leaders possessing such dimensions are capable of operating in crews that perform at relatively high levels after an initial adjustment period, but it may not be possible to maintain these levels over long periods of time. Over time, we would predict that individuals possessing these dimensions would have difficulty maintaining an effective crew process. If this is the case, we should be particularly concerned about crewmembers fitting this profile and participating in long-duration operations such as ship, submarine, or space station operations. The two-day time period of this simulation study was not sufficient to explore these limits, and it is important that research on longer durations be accomplished.

*Personality Theory Implications*

Foushee (1984) argued that flight simulation provides an ideal environment in which to conduct research meeting both basic and applied criteria. Since aircraft simulators provide high levels of both realism and experimental control, they make ideal laboratories for experimentation, and there are clearly fewer problems of generalizability to real world behavior. On the negative side, high-fidelity simulation demands a large investment, and sound use of these expensive facilities demands that proposed theories or models be conceptualized with real-world applications in mind, which suggests in turn that researchers collect evidence for the ecological validity of their theories prior to testing in a high-fidelity environment. The trade-off identified here argues for a program of research moving from lower-fidelity to increasingly high-fidelity environments.

We chose the selection criteria utilized in this study in large part because of the substantial body of real-world performance-relevant data collected by Spence, Helmreich, and their colleagues (Spence & Helmreich, 1983; Helmreich, 1982, 1986; Chidester, 1986, 1987). For example, Spence & Helmreich's (1983) measures of achievement motivation were shown to predict performance among academic scientists and engineers. Moreover, Helmreich (1986) found Instrumentality and Expressivity to be significantly correlated with check airman evaluations of individual pilot performance. This study has provided further evidence for the validity of Instrumentality and Expressivity as meaningful and important components of individual personality. As we consider the development of selection criteria for future aerospace operations, these dimensions appear to be strong candidates for representation.

We believe that studies of this type represent an important direction for personality research. Helmreich (1983) has argued that researchers have overly sterilized their work to the extent that it does not apply to real world phenomena. The failures of personality researchers to demonstrate strong links with important behavioral dimensions may be in large part the result of the "sterile" laboratory tradition so predominant in psychological research. The structuring of artificial tasks for laboratory experimentation, a process viewed as necessary for both control and assessment, may also tend to create artificial behavior that accounts for far more behavioral variation than the experimental variables themselves. This phenomenon may have been a factor in the problem-plagued search for personality predictors of performance in past research. Thus, high fidelity research environments may put us in a better position to resume our search.

*Summary*

The results of a highly realistic full-mission simulation highlighted the importance of the personality characteristics of pilots during routine and high workload flight segments. The personality profiles of leaders predicted how well crews dealt with significant operational events. IE+ led crews were consistently more effective than Ec- led crews, but I- led crews exhibited a familiarity effect. This change in performance over time among I- led crews raises questions that will be answered through analyses of process data from this study. New questions about the limits of familiarity must be addressed through future research. It seems clear that the impact of personality is significant and not of trivial magnitude. Selection approaches to optimizing crew effectiveness in aerospace operations should be more carefully considered. Personality theory and research should move towards a closer association with real-world performance.

## REFERENCES


1. Angyal, A. (1965). *Neurosis and treatment: A holistic theory.* New York: Wiley.

2. Bakan, D. (1971). *Slaughter of the innocents.* San Francisco: Jossey-Bass.

3. Blake, R. R., & Mouton, J. S. (1978). *The new managerial grid.* Houston: Gulf.

4. Chidester, T. R. (1986). Mood, Sleep, and Fatigue Effects in Flight Operations. Unpublished Doctoral Dissertation. The University of Texas at Austin.

5. Chidester, T. R. (1987). Selection for optimal crew performance: Relative impact of selection and training. In R. S. Jensen (Ed.),. *Proceedings of the Fourth International Symposium on Aviation Psychology,* Columbus, OH, 473-479.

6. Cohen, J. (1977). *Statistical power analysis for the behavioral sciences,* Revised edition. New York: Academic Press.

7. Fiedler, F. E. (1967). *A theory of leadership effectiveness.* New York: McGraw-Hill.

8. Foushee, H. C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist,* 39, 885-893.

9. Foushee, H. C., & Helmreich, R. L. (1988). Group interaction and flightcrew performance. In E. L. Wiener and D. C. Nagel (Eds.), *Human factors in aviation.* San Diego: Academic.

10. Foushee, H. C., Lauber, J. K., Baetge, M. M., & Acomb, D. B. (1986). Crew factors in flight operations III: The operational significance of exposure to short-haul air transport operations. *NASA Technical Memorandum 88322*, Moffett Field, CA: NASA-Ames Research Center.

11. Foushee, H. C., & Manos, K. L. (1981). Information transfer within the cockpit: Problems in intracockpit communications. In C. E. Billings and E. S. Cheaney (Eds.), Information transfer in the aviation system. *NASA Technical Paper 1875*. Moffett Field, CA: NASA-Ames Research Center.

12. Halpin, A. W. (1954). The leadership behavior and combat performance of airplane commanders. *Journal of Abnormal and Social Psychology*, 49, 19-22.

13. Helmreich, R. L. (1982). Pilot selection and training. Paper presented at the annual meeting of the American Psychological Association, Washington, D. C.

14. Helmreich, R. L. (1983). Applying psychology in outer space: Unfulfilled promises revisited. *American Psychologist, 38*, 445-450.

15. Helmreich, R. L. (1984). Cockpit management attitudes. *Human Factors*, 26, 583-589.

16. Helmreich, R. L. (1986). Studying flight crew behavior: A social psychologist encounters the real world. Invited address to the annual meeting of the American Psychological Association, Washington, D. C.

17. Helmreich, R. L., Sawin, L. L., & Carsrud, A. L. (1986). The honeymoon effect in job performance: Delayed predictive power of achievement motivation. *Journal of Applied Psychology*, 71, 1085-1088.

18. Kanki, B. G., Lozito, S. C., & Foushee, H. C. (1988, in press). Communication Indexes of Crew Coordination. *Aviation, Space, and Environmental Medicine.*

19. Lauber, J. K., & Foushee, H. C. (1981). Guidelines for line-oriented flight training. *NASA Conference Publication 2184*, Moffett Field, CA: NASA-Ames Research Center.

20. Melton, A. W. (Ed.) (1947). *The AAF qualifying exam, Air Forces Aviation Psychology Program Research Reports*, Number 6. Washington, D. C.: U. S. Government Printing Office.

21. McGrath, J. E. (1964). Social psychology: A brief introduction. New York: Holt.

22. McGrath, J. E. (1984). *Groups: Interaction and performance.* Englewood Cliffs, N. J.: Prentice-Hall.

23. Norusis, M. J. (1988). Chapter 3: Cluster Analysis. In M. J. Norusis (Ed.),. *SPSS/PC+ Advanced Statistics V2.0.* Chicago: SPSS Inc.

24. Orlady, H. W., & Foushee, H. C. (1987). Proceedings of the NASA/MAC workshop on cockpit resource management. *NASA Conference Publication 2455*, Moffett Field, CA: NASA-Ames Research Center.

25. Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, 85, 185-193.

26. Ruffell-Smith, H. P. (1979). A simulator study of the interaction of pilot workload with errors, vigilance, and decisions. *NASA Technical Memorandum 78482*. Moffett Field, CA: NASA-Ames Research Center.

27. Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* Austin, TX: University of Texas Press.

28. Spence, J. T., & Helmreich, R. L. (1983). Achievement-related motives and behavior. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches.* New York: W. H. Freeman & Co.

# Reactions to Emergency Situations in Actual and Simulated Flight

Glenn F. Wilson
June Skelly
Bradley Purvis

Human Engineering Division
Armstrong Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, OH 45433-6573
U.S.A.

## SUMMARY

Four emergency situations occurred during inflight and simulated air-to-ground training missions. Heart rate data were recorded from the pilots as part of a study designed to determine the effects of mission segment and flight position. Fifty percent increases in heart rate were found to occur only during actual flight but not during simulated flight emergencies. Heart rate variability decreased in all cases but to a greater extent during the inflight emergencies.

## INTRODUCTION

Flight produces both physiological and psychological stress on crew members. While this has been true since before the first flight it is especially true with modern aircraft and in particular in high performance fighter aircraft. The effects of these stressors is usually unmeasured. Laboratory and simulator experiments coupled with modeling efforts provide data which is used to make judgments about human tolerances to the various stresses of the flight environment. There have been a number of projects which have collected inflight physiological data, usually cardiac events (1,2,3,4,5,6,7). This can provide very useful information for the flight surgeon and human factors engineer. However, it has been observed that the stresses of the flight environment are very different from those of the simulator and further, that the active role of piloting an aircraft produces yet different stresses on the pilot. Actual piloting vs merely being in the cockpit and piloting vs being under autopilot control produce very different physiological responses (5).

Emergency situations place crew members under additional stress and a great deal of training time is spent teaching crew members how to recognize and appropriately react to the many emergencies that can befall modern aviators. Since emergencies are rare, but may lead to catastrophic results, they are of great interest to the aerospace community. The usual evidence available from inflight emergencies is in the form of subjective report, damage evidence and eye witness reports. Physiological reactions to inflight emergencies are difficult to record since one can not predict when they will occur and pilots are not routinely instrumented for physiological data recording so that when they do happen the physiological data are lost. If the same differences exist in physiological responses between simulated and actual emergencies that exist between different aspects of normal flight, then the responses to simulated emergencies are likely to only approximate those responses to actual emergencies. One must keep in mind the oft quoted statement "no one has ever died in a simulator".

We collected physiological data from four pilots during emergency situations during the conduct of a study on the effects of mission segment and flight formation position. Two of these events occurred during actual flight and two occurred during simulated flight. These serendipitous events provide us with the opportunity to investigate the physiological events which accompanied these emergencies and to compare them with other segments of normal flight.

The larger study was designed to evaluate the physiological effects of mission segment and flight formation position on pilot heart rate and eye blink activity. Air-to-ground training missions were flown with each of the pilots flying the lead and wing positions and also in a simulated version of the mission. This permitted us to determine the effects of the various segments of the mission on the physiological responses, and to also compare the effects of flying lead or wing in actual and simulated missions. Heart rate and eye blinks were significantly affected by both of these variables. These data will not be reported here, but are available in other proceedings (8,9). The typical training mission took about 90 min. and included a high altitude flight from home base to a low level segment; flying the low level segment, several turns and navigation updates, a weapons delivery segment and high altitude cruise back to base.

The purpose of this paper is to describe the physiological changes resulting from the emergency events and also to compare these responses with those from other segments of the missions. The emergency responses themselves are of interest as is their relationship to the larger context of normal flight.

## METHODS

Subjects. Eight A7 pilots participated in this study. All were qualified to fly both lead and wing positions in multi-ship formations. The four pilots whose data will be discussed here had an average of 833 hours flying the A7.

Aircraft and Simulator. A-7D single seat aircraft were used in this project. They were equipped with cannons, missiles and gravity weapons. The simulator was a high fidelity A-7 Weapons Systems Trainer equipped with 135 degree vision system, full instrumentation emulation, heads up display, radar, and six degrees of motion. The simulator mission scenario was as close to the actual aircraft flight scenario as possible in order to maintain a realistic comparison.

Physiological Data. Each pilot was instrumented with 3M Red Dot disposable Ag-Ag/Cl electrodes to pick-up cardiac activity and were placed over the manubrium and over the 5th rib on the left thorax with a common ground electrode placed over the 7th rib on the right side. Del Mar Avionics Model 509 Neurorecorders were used to record the data and were worn in the leg pocket of the G-suit. Electrooculograms were also recorded but will not be reported here.

Audio Recorder. The pilots were equipped with a small audio tape recorder which was connected to the aircraft communication system. This recorder provided a continuous auditory record of each mission so that significant mission events could be located in time and correlated with the time channel of the physiological recorders.

Data Analysis. The mission audio tapes were transcribed and the time of occurrence of the emergencies and other events of interest were noted. Two minutes of data were analyzed for all segments. The ECG data were filtered between 4 Hz and 50 Hz to simplify R wave detection. These data were then digitized and analyzed using AAMRL's Neuropsychological Workload Test Battery (10), which provided heart rate and inter-beat-interval (IBI) data. The cardiac data will be presented in two ways, mean rate for successive 10 sec epochs and as a continuous cardiotach record. The cardiotach record shows the IBI data and also makes it possible to visualize the variability of these data. High variability in the IBI intervals has been associated with low workload levels while low or no variability has been shown to accompany high stress or high workload events. However, other conditions can also reduce variability such as speech, physical effort, age, etc. Typical heart rate variability (HRV) measures require several minutes of stable data, and they are not suitable to quantify transient events. Therefore, visual inspection of the IBI traces will be used to describe the HRV of the present data.

Besides the emergency data, three other segments will be described. In order to provide a broad comparison, normal and high stress flight segments and the pre-flight briefing segment will be included. The level flight, cruise segment, will serve as an inflight baseline since workload is relatively low. The two minute segment prior to brake release at takeoff provided high stress data for all pilots. The briefing segment was taken during the middle of the approximately one hour pre-flight briefing.

## RESULTS

The four emergency events will be described separately, the flight data first followed by the simulator data. Each segment will be described and then compared to data collected during the non-emergency segments of each mission.

Inflight Data. The first incident involved a bird strike during the low level portion of the mission. As soon as the pilot heard the strike he alerted the other members of the flight, in which he was flying wing, and immediately gained altitude. Another ship then flew around his aircraft to visually assess the damage to the ship. These two ships then proceeded back to base and a normal landing. There was no damage to the aircraft, apparently only the bird suffered severe damage. The pilot's heart rate (HR) increased 53% following the strike compared to the pre-strike level, from 85 bpm to 130 bpm. The cardiac data are shown in Figure 1 with the 10 sec averages presented above the continuous IBI trace. With this pilot, the maximum rate was reached after about 30 secs and this accelerated rate returned to pre-strike levels in about 80 sec. The mean HR for the entire two minute epoch was 103.6 bpm. The cardiotach record shows that the normal rhythmic nature of the pre-strike cardiac activity, HRV, was lost shortly after the strike and remained flat for approximately 60 to 70 secs following the strike.

The HR and IBIs associated with the cruise segment of the mission prior to the low level portion are depicted in Figure 2 and show that the levels of cardiac activity are consistent and demonstrate a large amount of HRV. The overall HR mean was 85.8 bpm which is the same as the pre-strike levels recorded later in the mission. The HR during this segment was consistent. The two minute mean for the pre-takeoff data was 91.0 bpm and displayed HRV. The 10 sec HR means ranged from 81.8 to 97.3 bpm during this segment. The HR and IBIs are presented in Figure 3.

The IBIs and HR during the pre-flight briefing are shown in Figure 4. The overall mean HR was 72.3 bpm and again HRV is evident and the HR is consistent across this time period. These records are in contrast to the bird strike segment in both HR changes and the amount of HRV. The two minute mean HRs for all four of the segments are listed in Table 1 for all four pilots.

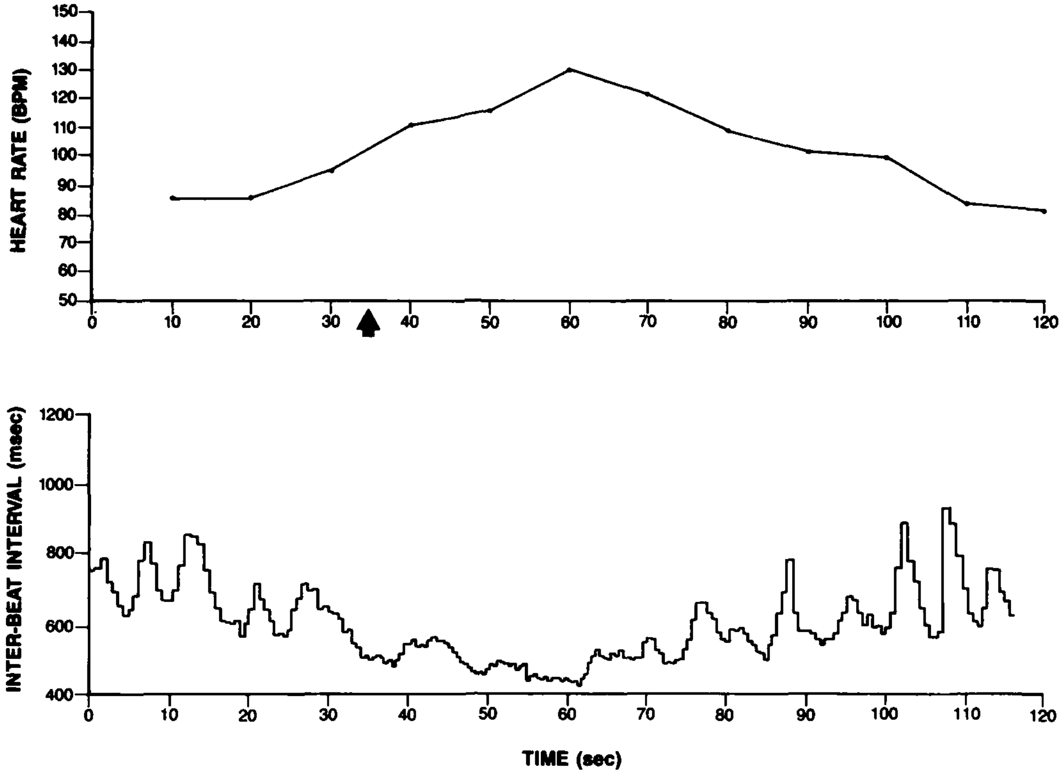



Figure 1. Ten second HR Means and IBIs for F1 during the emergency segment. The arrow denotes the time of the bird strike.

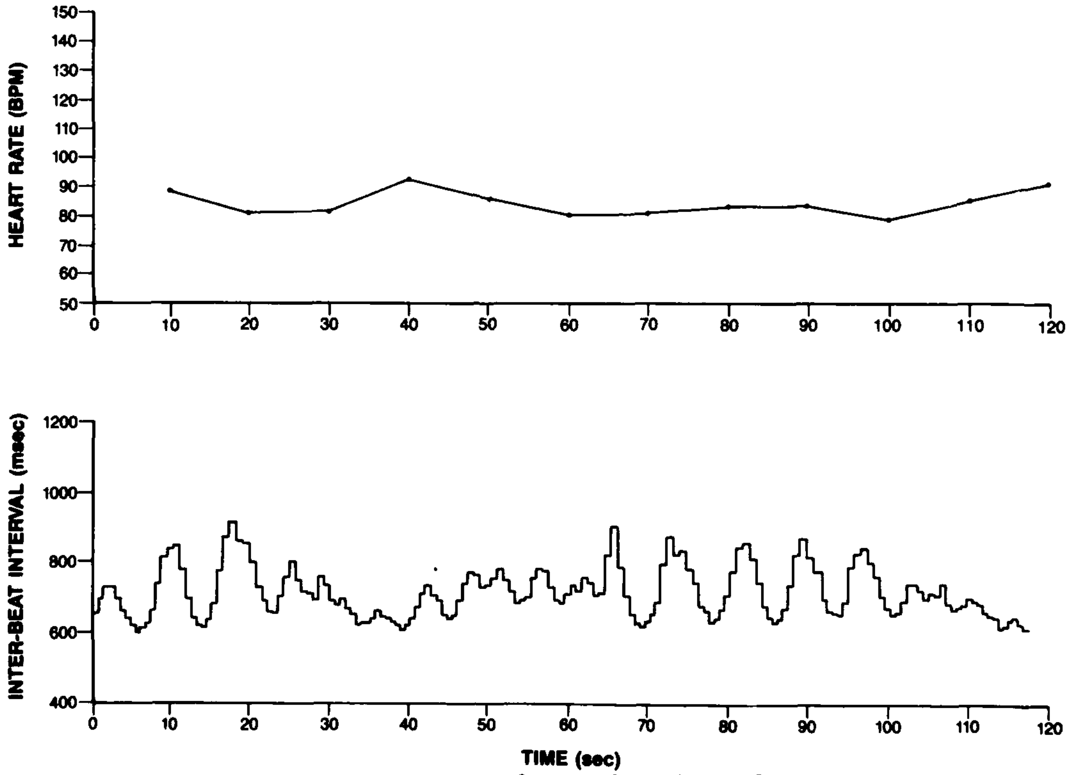



Figure 2. Inflight cruise segment HRs and IBIs for pilot F1.

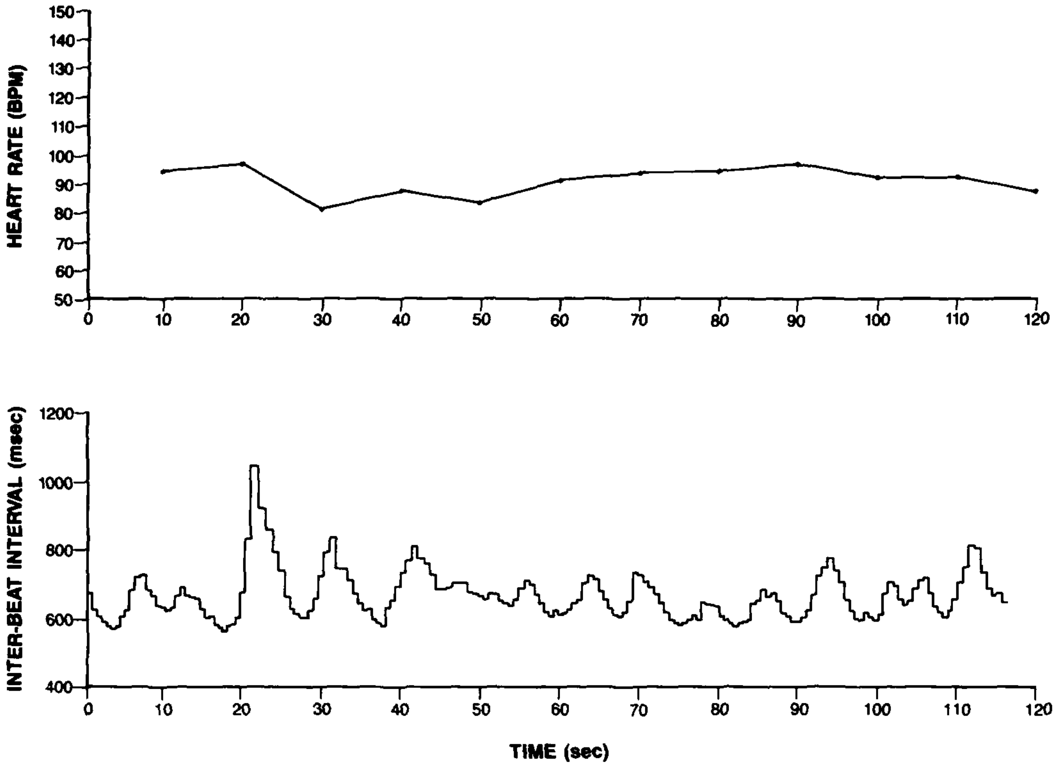



Figure 3. Pre-takeoff HRs and IBIs for pilot F1.

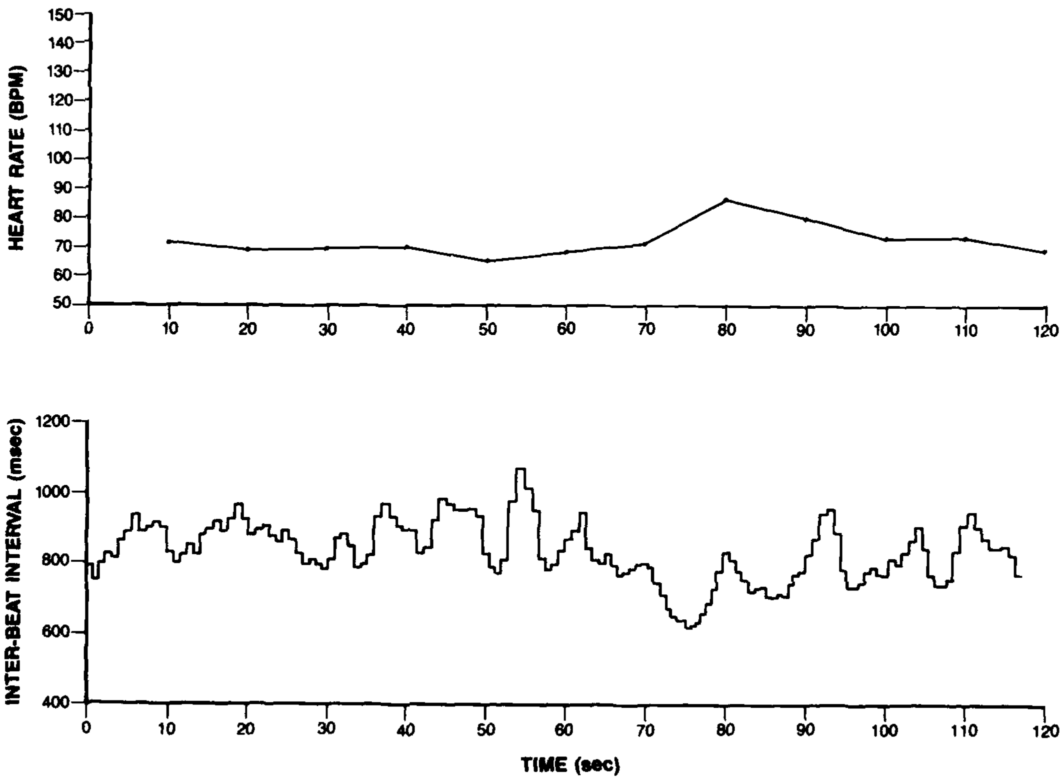



Figure 4. Briefing HR's and IBI's for pilot F1.

Table I.

Mean HR for the 120 sec epochs for the two inflight (F1 and F2) and the two simulator (S1 and S2) pilots. Event is the emergency segment, Pre T/O is pre-takeoff, cruise is the inflight level segment and Brief is the pre-flight briefing data.

| | Event | Pre T/O | Cruise | Brief |
|---|---|---|---|---|
| F1 | 103.6 | 91.0 | 85.8 | 72.3 |
| F2 | 86.6 | 81.4 | 64.6 | 59.2 |
| S1 | 89.8 | 92.0 | 93.5 | 73.8 |
| S2 | 96.5 | 85.0 | 87.8 | 81.3 |

The second inflight emergency, which also occurred during the low level segment, involved a situation in which one of the wing pilots felt that the lead aircraft was about to fly into his path. An evasive maneuver was performed followed by reforming and continuation of the mission. This event was associated with an increase in HR of 51%, from 70 to 106 bpm, which occurred within a matter of seconds. As can be seen in Figure 5, the HR remained at that level for 20 sec before it started to decline. The return to pre-emergency levels took a further 60 sec. The mean HR for this two minute epoch was 86.6 bpm. This pilot did not exhibit a great deal of HRV in any of his records making it difficult to assess any effects on the variability.

The flying baseline, cruise segment, is depicted in Figure 6. The HR and HRV are extremely stable during this portion of the mission which was characteristic of this pilot's recordings. There is little evidence of HRV from these records and quite stable activity levels were found with a mean of 64.6 bpm during this 120 sec epoch. The pre-takeoff HR and IBI data are presented in Figure 7 and show a step increase after the first 30 - 40 sec. The data then remain stable to the end of the two min epoch. The HR ranged from a 10 sec mean low of 66.3 to a high of 96.0 bpm. Prior to the increase the HR was in the high 60 bpm range and then was consistently in the 80 bpm range. Little HRV is exhibited in these records. Data recorded during this pilot's briefing are depicted in Figure 8. The HR is a bit more variable, mean of 59.2 bpm, with some HRV evident but still representing a stable record. The means for each of the two minute segments are listed in Table 1.

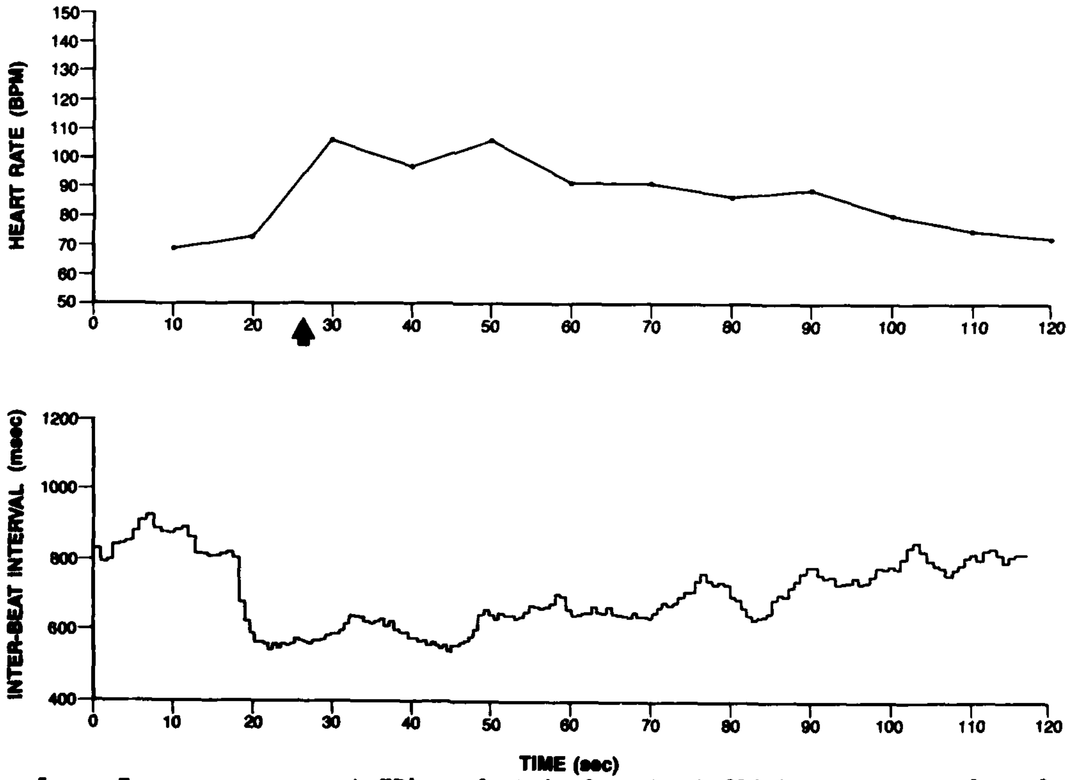



Figure 5. Emergency segment HR's and IBI's for the inflight emergency data for pilot F2. The arrow denotes the time of the occurrence.

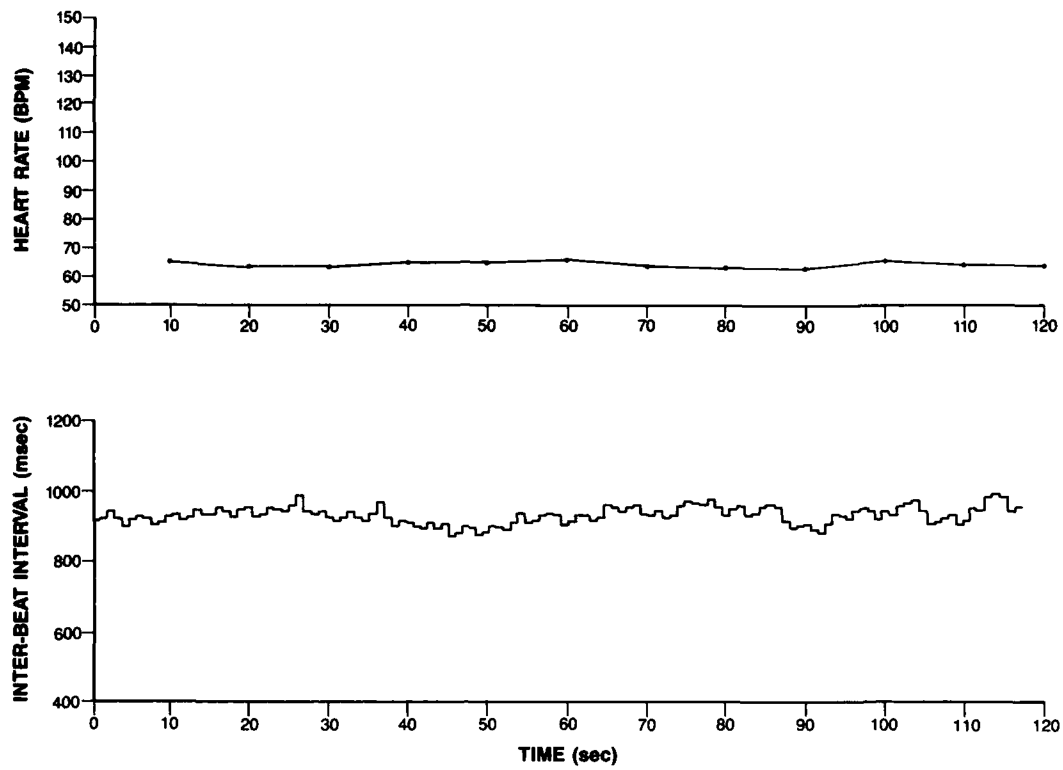



Figure 6.    Inflight cruise segment HR's and IBI's for pilot F2.

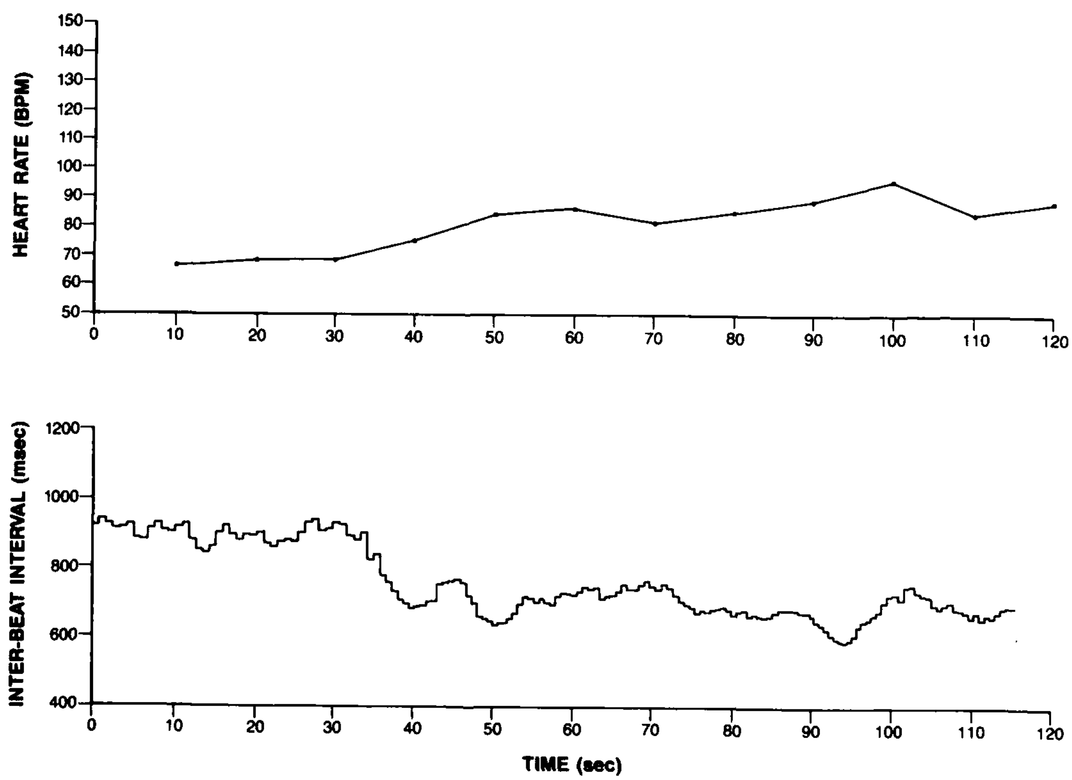



Figure 7.    Pre-takeoff HR's and IBI's for pilot F2.

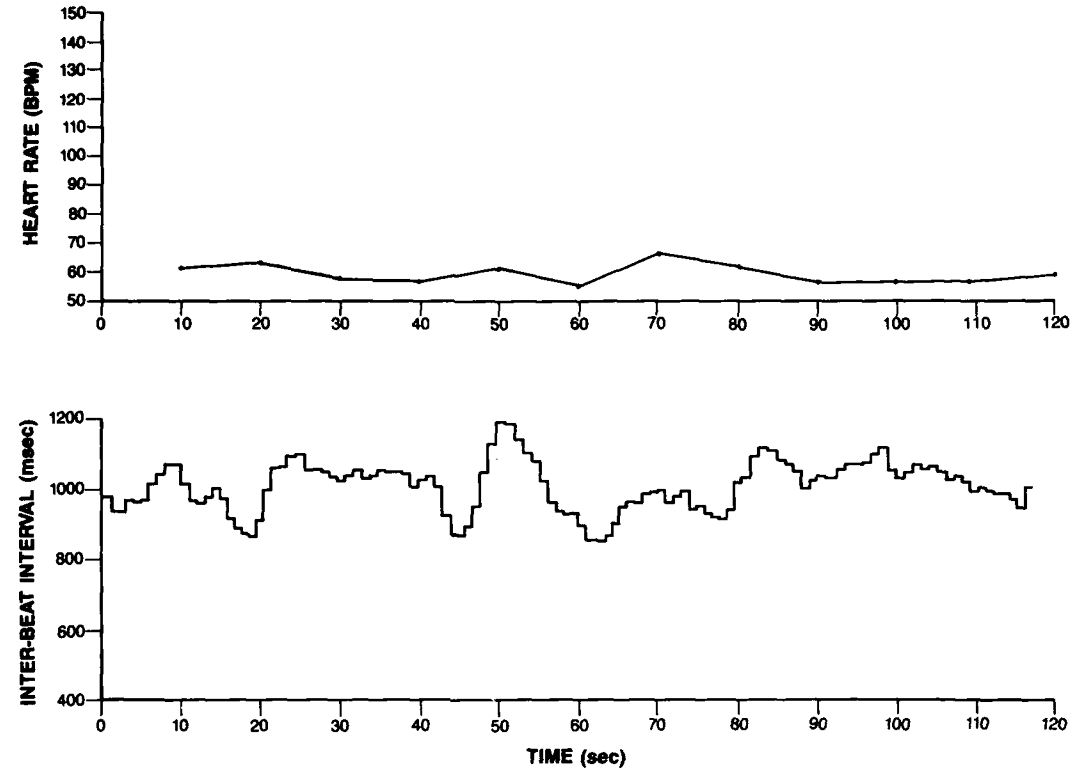



Figure 8. Briefing HR's and IBI's for pilot F2.

Simulator Data. Both of the simulator emergencies were crashes that occurred during a portion of the weapons delivery segment that was designed to produce stress in the pilots. It involved numerous threats not included in the inflight missions but which were included only in the simulation to determine the effects of higher levels of threat upon pilot performance. Data from the first simulator mission are presented in Figure 9. A small increase in the HR was evident following the "crash". This 11 bpm acceleration represented a 13% increase in the HR from the lowest HR of 84 bpm, at the crash, to the highest 10 sec average rate of 95 bpm. The overall mean HR for the 120 sec of data was 89.8 bpm. There was a decrease in the HRV, as seen in the IBI data, following the "crash". The variability does return within about 20 sec.

The cruise segment for this pilot is shown in Figure 10. The overall mean was 93.5 bpm with a range from 85.7 to 101.5 bpm which is larger than that of the "crash" data. There is HRV as seen in this pilot's IBI data. The pre-takeoff data had an overall mean of 92.0 bpm and were fairly stable with a range from 87.0 to 101.5 bpm. Consistent HRV is evident in the IBI graph of Figure 11. The briefing segment, Figure 12, shows that there is definite HRV with a HR mean of 73.8 bpm and a 10 sec mean range from 65.1 to 78.9 bpm. The HR for this segment is lower than either of the flight segments; cruise or "crash".

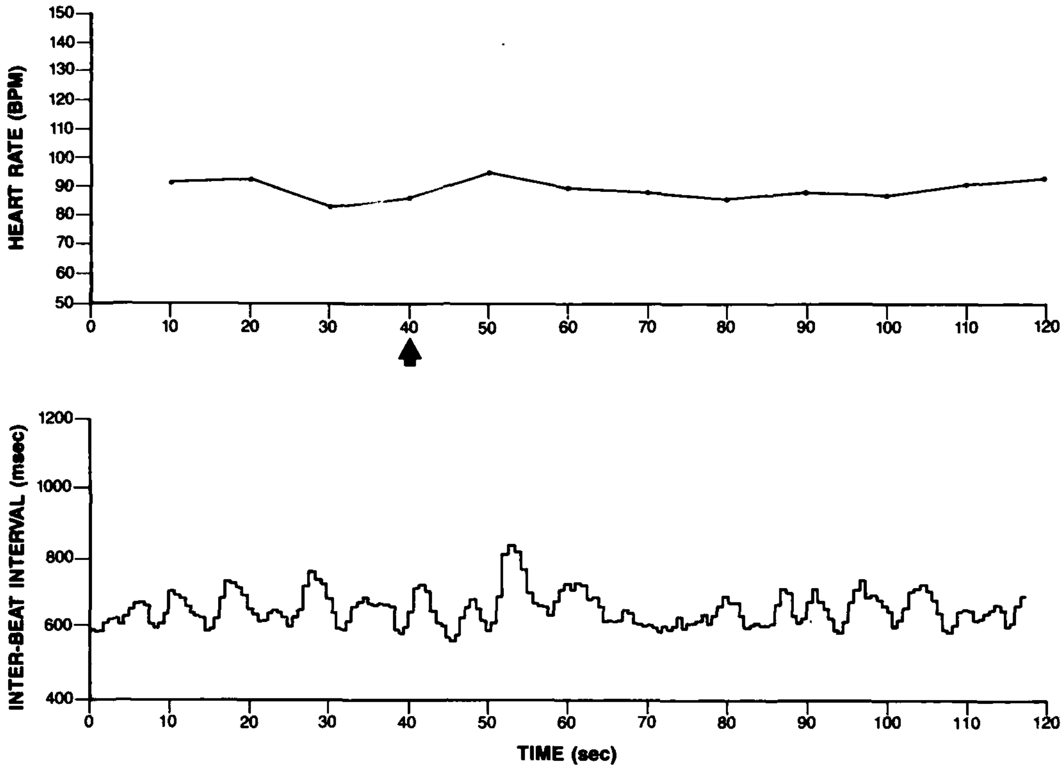



Figure 9. Emergency segment HR's and IBI's for simulator "crash" for pilot S1. The arrow denotes the time of the "crash".

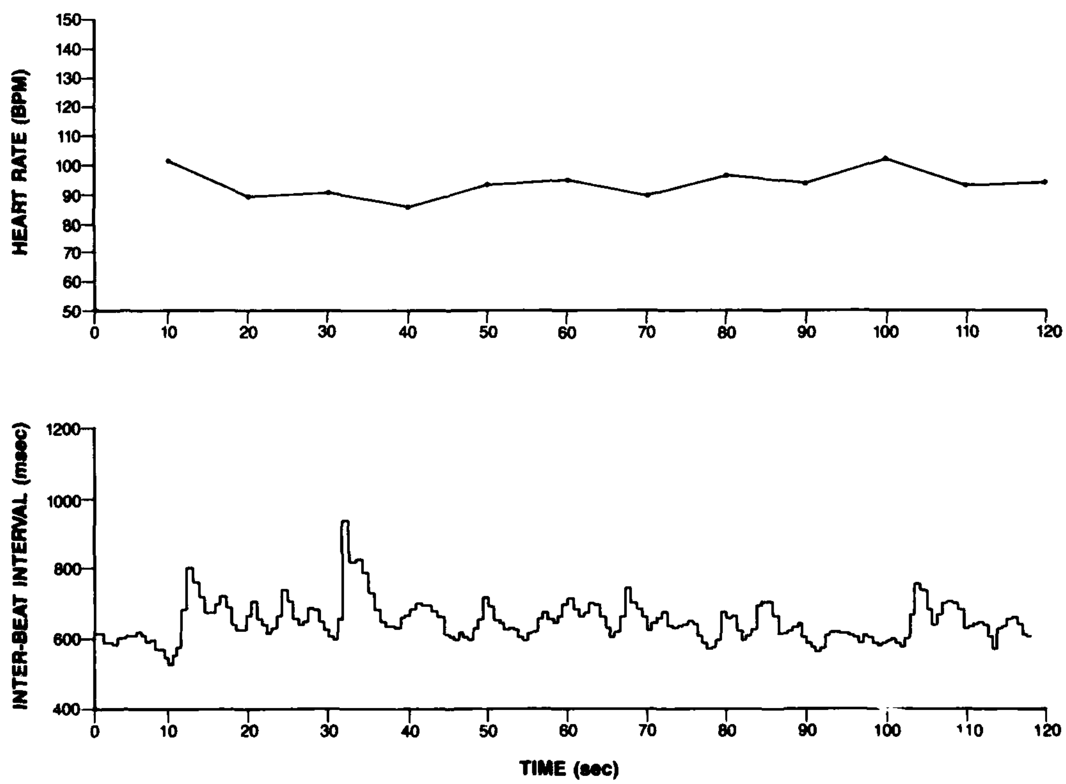



Figure 10. Simulator cruise segment HR's and IBI's for pilot S1.

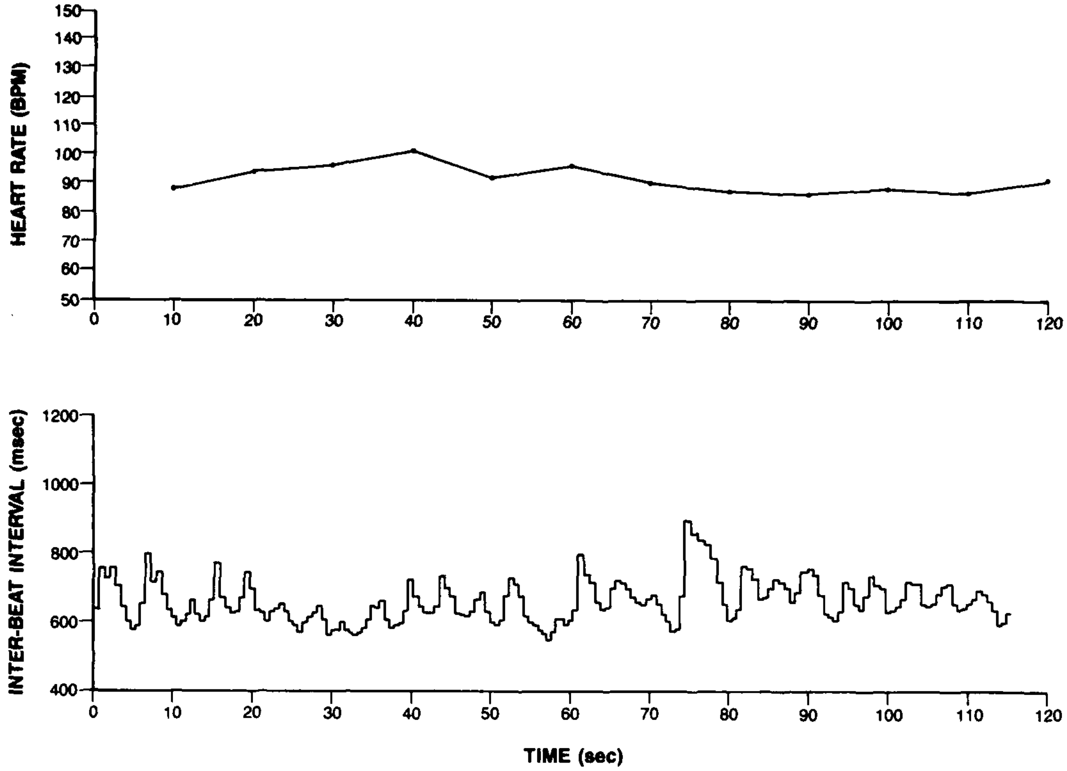



Figure 11.   Simulator pre-takeoff HR's and IBI's for pilot S1.

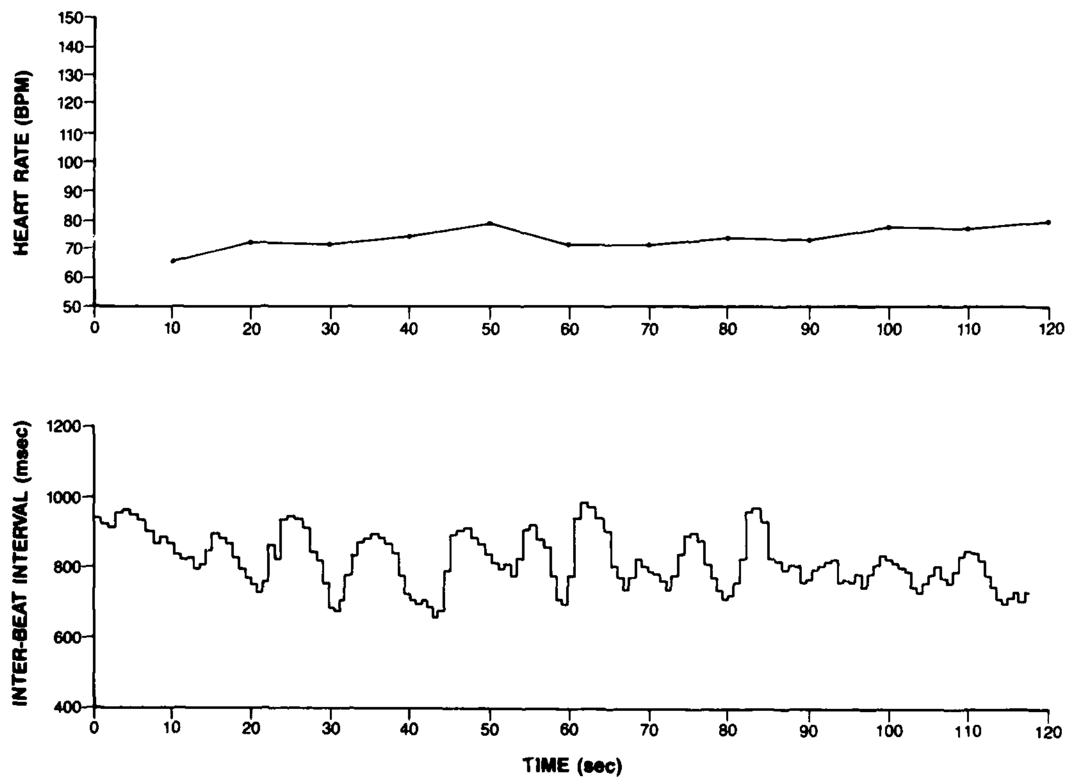



Figure 12.   Briefing HR's and IBI's for pilot S1.

The fourth emergency event also occurred during the end of the weapons delivery portion of the simulated mission. As can be seen in Figure 13, there is a drop in HR midway through the 120 sec segment. HRV is evident at the beginning and ending of this segment but is suppressed just before and during the "crash". The mean HR for this entire segment is 96.5 bpm. The inflight cruise segment, Figure 14, shows a variable HR, ranging from 81.5 to 92.3 bpm with a mean of 87.8 bpm. The pre-takeoff data are depicted in Figure 15 and show consistent HR levels with a mean of 85.0 bpm and a 10 sec epoch range from 81.2 to 92.9 bpm. The IBI graph shows HRV, especially during the middle portion of this segment. The briefing data are shown in Figure 16. There is a noticeable increase in HR at 50 sec followed by a return to a stable level. *The IBIs show some variability early and later in the segment with little* evident during the middle segments. The mean HR was 81.3 bpm with a range from 73.6 to 95.4 bpm for the 10 sec averages.

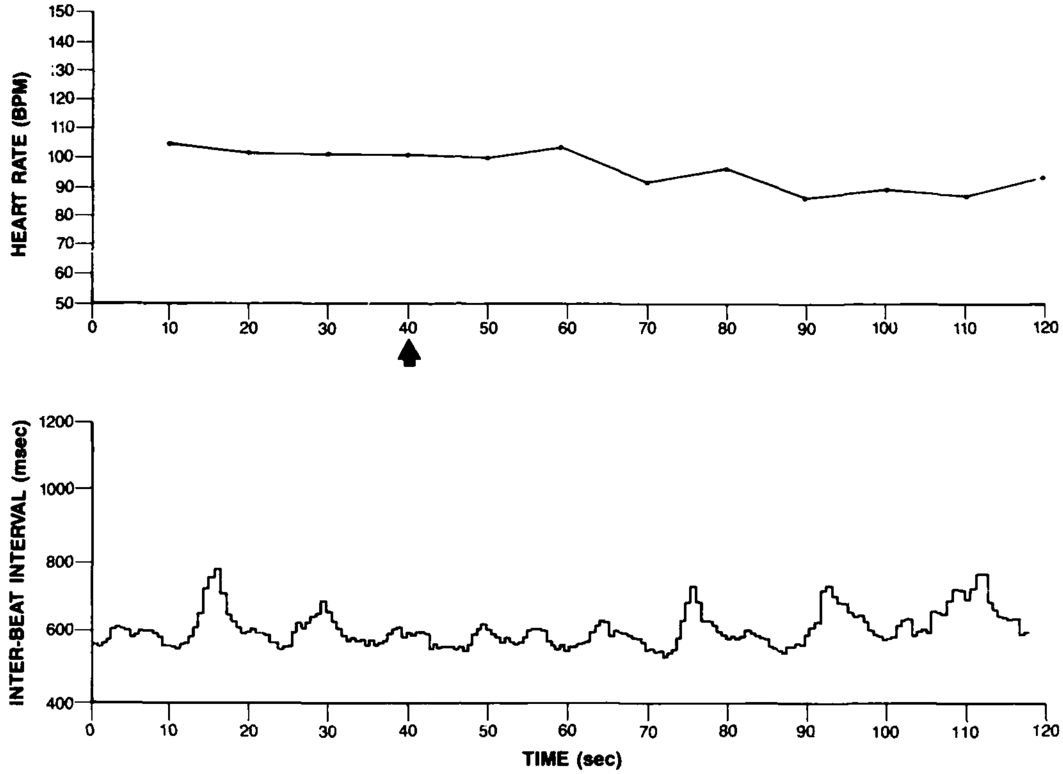



Figure 13. Emergency segment HR's and IBI's for simulator "crash" for pilot S2. The arrow denotes the time of the "crash".

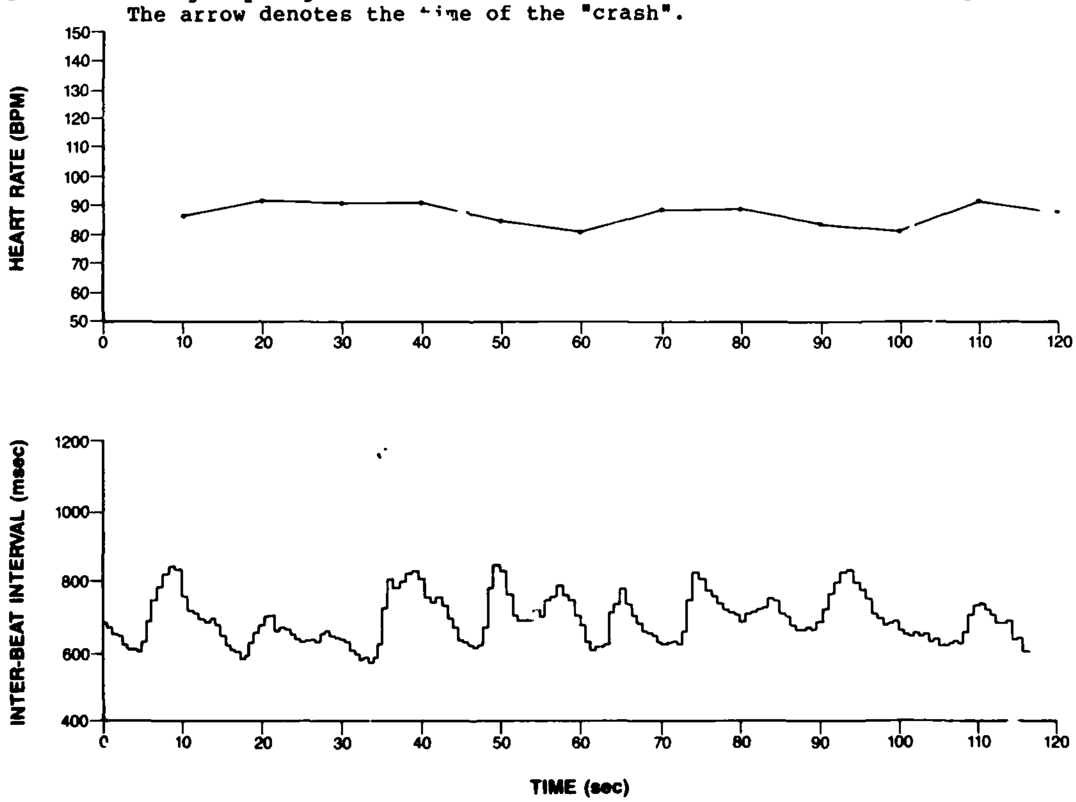



Figure 14. Simulator cruise segment HR's and IBI's for pilot S2.

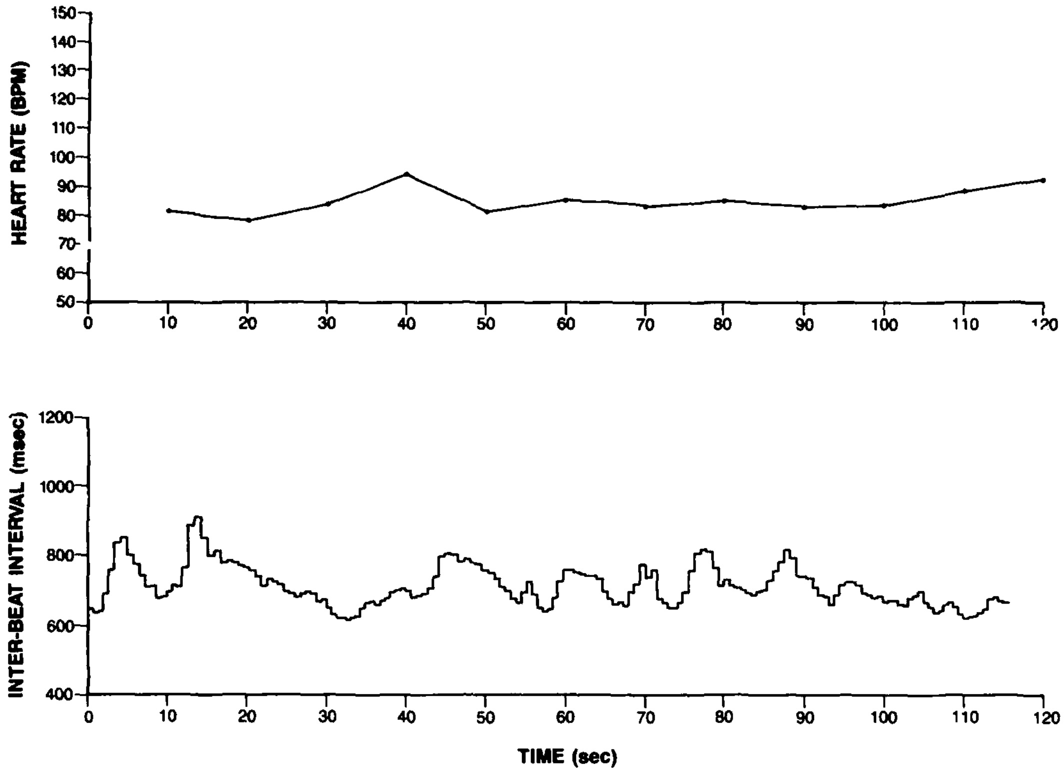



Figure 15. Simulator pre-takeoff HR's and IBI's for pilot S2.

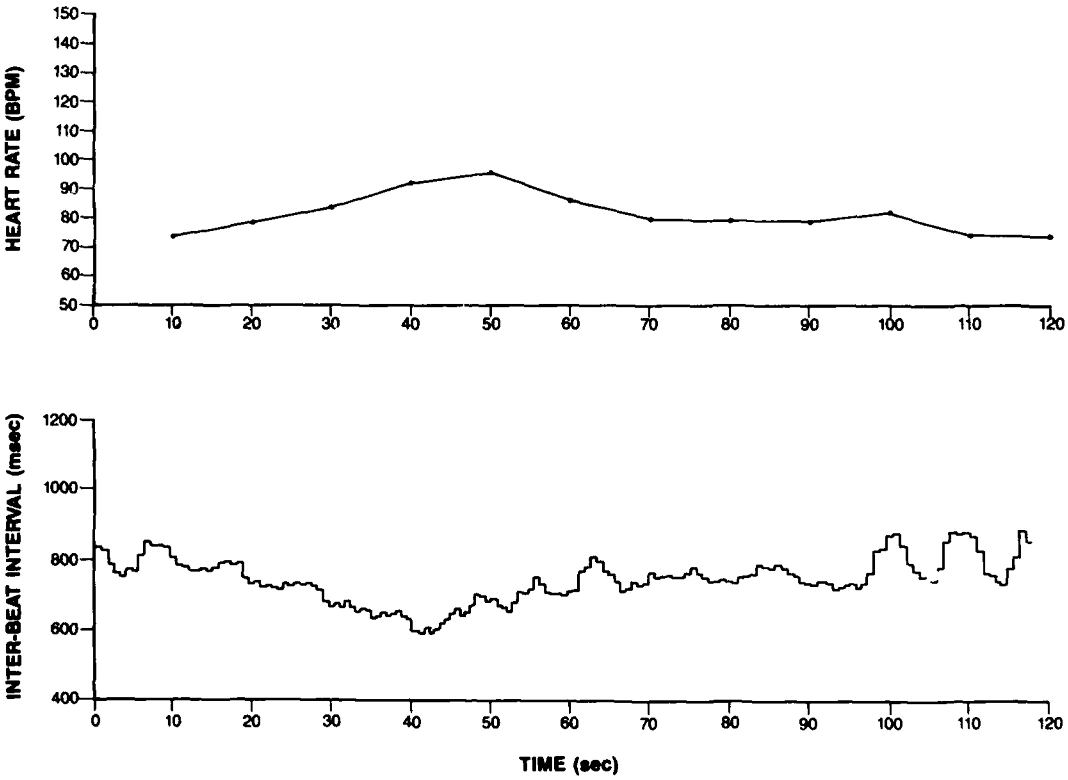



Figure 16. Briefing HR's and IBI's for pilot S2.

# DISCUSSION

This unique set of data provides the opportunity to see the effects of inflight emergencies on pilot HR. Further, it makes possible the comparison of the emergency data with other inflight segments of high and low stress. The simulator data provide the opportunity to make the same comparisons and also contrast these effects with the inflight data.

Both of the inflight emergencies were followed by a greater than 50% increase in HR. This is interesting since these pilots' pre-emergency HRs were 20 bpm different (e.g. 65 and 86 bpm). Even though the rate of change in HR to the maximum was different for the two pilots, 30 sec vs 2 - 5 sec, in both cases, the return to the pre-emergency levels took about 60 - 80 sec. In light of the nature of the events precipitating these HR changes it seems remarkable that the return was so rapid. This could well be due to the pilots' emergency training, their substantial experience and confidence in their ability to control the situation. These responses to emergencies were compared to the two other inflight segments, one representative of normal low workload and low stress (cruise) and the other representing the highest workload and stress segment as determined by HR (pre-takeoff). The peak HR and overall mean HR was higher for the emergency data and were characterized by a sudden transient increase in HR. The pre-takeoff segments represented the highest workload for normal flight segments, but produced lower HR without the transient increase seen in the emergency data. The cruise segment, in comparison, was characterized by relatively stable and lower HRs with high HRV in the one pilot who displayed HRV. The briefing segment produced the lowest overall HR in both pilots which is not surprising due to the overall low stress involved in this routine ground based segment. It seems that the emergency events are characterized by a sudden large increase in HR and a decrease in HRV that are greater than responses to the highest workload events in flight. In contrast, the simulator emergency HRs were not very different from the other "inflight" or briefing segments. While there were small differences across the four segments presented here, they were not consistent for both pilots. There does seem to be a decline in the HRV during the emergencies that is not evident in the other segments. The low stress of the simulated missions seems to have little effect on the pilots' cardiac activity and the simulator "crashes" do not produce the magnitude of change, if any at all, that was characteristic of the inflight emergencies. While the HRs did not show marked increases, the HRV did decrease following the crashes. It is possible that HRV is more sensitive to workload type effects, while HR is more of a stress indicator.

For these two pilots and the larger group in the original study, the simulator HRs were significantly lower than inflight HRs, but were not statistically different across mission segments. However, the inflight HRs were significantly affected by flight segment. Overall, the HR data obtained in the simulator are more uniform, no doubt due to the pilots' perception of a low stress environment.

The distinct nature of the inflight emergency data following each emergency suggests that it may be possible to automatically search cardiac data for this type of event. The magnitude and transient nature of the HR responses and change in the HRV could provide a signal to the system that a very stressful event had occurred and this information could be used by the system, in conjunction with indicators from other elements, to make appropriate corrections. Baseline data for each pilot or system operator would be required so that the system could recognize the abnormal perturbation in the person's physiological responses. Even if the system was otherwise alerted by information from other subsystems, the stress response from the pilot would be quite meaningful with regard to the appropriate response required of the system. The presence of a "stress" response from the pilot indicates that they are aware of the situation and a continued stress response would indicate that system intervention was appropriate. If a stress response was not present but other information indicated that a severe problem existed then the pilot must be alerted and/or a determination made whether the pilot was still capable of responding.

# REFERENCES


1. Comens, P., Reed, D. and Mette, M. Physiologic responses of pilots flying high-performance aircraft. Aviation, Space, and Environmental Medicine, 1987, 58, 205-210.

2. Billings, C. E., Gerke, R. J, Chase, R. C. and Eggspuehler, J.J. Stress and strain in student helicopter pilots. Aerospace Medicine, 1973, 44, 1031-1035.

3. Roman, J., Older, H and Jones, W. L. Flight research program: VII. Medical monitoring of navy carrier pilots in combat. Aerospace Medicine, 1967, 38, 133-139.

4. Roman, J., Perry, J.J., Carpenter, L. R. and Shaiban, A. A. Flight research program: VI. Heart rate and landing error in restricted field of view landings. Aerospace Medicine, 1967, 38, 128-132.

5. Roscoe, A. H. Use of pilot heart rate measurement in flight evaluation. Aviation, Space, and Environmental Medicine, 1976 47, 86-90.

6. Roscoe, A. H. Assessing pilot workload in flight. In: Conference Proceedings No. 373, Flight Test Techniques, AGARD, Paris, 1984.

7. Ruffell Smith, H. P. Heart rate of pilots flying aircraft on scheduled airline routes. Aerospace Medicine, 1967, 38, 1117-1119.

8. Skelly, J. J., Purvis, B., and Wilson, G. Fighter pilot performance during airborne and simulator missions: Physiological comparisons. In, Electric and Magnetic Activity of the Central Nervous System: Research and Clinical Applications in Aerospace Medicine, AGARD Symposium, Trondheim, Norway, 1987, 23-1 to 23-15.

9. Wilson, G. F., Purvis, B., Skelly, J., Fullenkamp, P., and Davis, I. Physiological data used to measure pilot workload in actual flight and simulator conditions. Proceedings of the Human Factors Society, 1987, 779-783.

10. Wilson, G. F. and O'Donnell, R. D. Physiological Measures of Mental Workload. Chapter in Hancock, P. and Meshkati (Eds) Human Mental Workload, Amsterdam, The Netherlands: Elsevier, 1988.

## DISCUSSION

FOUSHEE: Could you tell us a little bit about how you did the simulation segments? I will tell you why I'm asking. There is a well known phenomenon among pilots, which we call the 'simulator syndrome'. Pilots who spend so much time undergoing training in a simulator typically get into the simulator with a sort of mental or cognitive set in which they have some expectation about the types of exercises they are going to be required to do. They become programmed, they know what to expect and they know exactly what is going to happen to them. That, of course, is very different from the real world. One of the things that we have found, and we have spent a lot of time looking at, is crew behaviour in full mission simulation exercises of more realistic types. I will have to oversimplify our findings; essentially, we tend to see more normal, more stress-type of reactions when the events are unexpected, especially when they occur after some period when nothing has been happening. Could you comment?

WILSON: Yes, the scenario was made to be as close to the actual training mission as possible, given the constraints of the simulator. These pilots did not fly the simulator very often. I'm trying to remember if it was only one to three times a year they actually flew in a simulator. They had to go to a different base to fly it. They knew it would probably be something similar to an actual mission because we told them basically what we were looking for. However, they did not know exactly what was going to happen or if we were going to throw anything particularly strange at them. The crashes occurred after the standard mission was over. We had a lot of SAM sites and things like that. It was a little novel to them and they certainly did not do it every day or every month.

PSIMENOS: When one is under stress the normally expected reaction is to have tachycardia, but during an actual flight did you observe any arhythmias that might lead to some organic trouble?

WILSON: We did not really look for them. Most of the analysis was done by computer, so we did not actually visually inspect or look for arhythmias or anything like that.

PSIMENOS: We tried to do the same. We used a heart monitor during actual flights in fighter aircraft; we were looking for arhythmias because cardiologically these are very important.

WILSON: Especially during the G-segments. We would be interested to know if arhythmias occurred, but we did not look for them.

THACKRAY: The heart rate increase that you got in flight, which was the result, I believe, of the bird impact, is very similar to the kind of response that I will be reporting on Thursday in terms of startle. Essentially, you had a traumatic, rapid onset emergency event, which gives a rather precipitous heart rate rise which, in our experiments, recovers within about one minute. Although I am not too clear how you simulated your emergencies, I assume that the simulation emergency was of the slower onset type rather than the precipitous type of event that occurred during flight. Is that correct?

WILSON: Yes, after our subjects had completed the normal mission we asked them to go back to the weapon delivery area where there were more numerous threats which were very difficult to evade. We could do this in the simulator but not, of course, in the real missions. Our pilots were flying through the high threat area for several minutes. It was not like a bird strike, which is a sudden event, but rather one that was going on for a while. What happened was these two fellows, in trying to evade, got too close to the ground and crashed.

THACKRAY: So the profile of the emergency situation to some extent determines the heart rate response. An emergency that is the result of a high work load situation, imposed upon individuals, gives a quite different heart rate response to one that occurs as the result of a very traumatic, rapid onset, emergency.

WILSON: Yes, I think the interesting thing is that when the crashes occurred heart rate was not that different from any other portion of that simulator segment, although you might want to claim that the workload was higher. The mean heart rates, however, were not that different from the previous 2 min segments that we looked at.

STRONGIN: I recently assisted in an investigation on an F-16 accident that is relevant to the topic we are discussing, i.e. state-dependent learning. We had an aircrew member who lost an engine on take-off and made a decision that was not necessarily directly in accord with standard emergency procedures but it was what most pilots would do. The issue was: How valid were his emergency procedures training in the simulator? Because the emergency he encountered could not be trained on a routine take-off it would have to be trained at altitude. Given what we know about state-dependent learning, it suggests that we might need to pay more attention to training emergency procedures, particularly sticking to "bold face" early in a pilot's career when he experiences anxiety both in the simulator and in actual flight. The issue of state-dependent learning and the probability of eliciting the most likely response is one that we have to consider in accident investigations. It is obviously not sufficent to say that you learn the same way in a fixed base simulator as you do in a real operational situation.

URSIN: Regarding the last point, I think it takes more than just heart rate to describe the physiological status that may or may not produce state-dependent learning in a human. Whether or not that does indeed exist we don't know, but it takes a lot of cues to have the phenomenon. A brief heart rate increase for 30 or 60 s, such as you describe, was shown very clearly in the data that we published in 1974; so old now that the reference does not appear on computer search. What we called "additional heart rate" was found in even very experienced pilots every time they landed their aircraft. It was not due to the workload because the workload was about the same in the first and second pilot. It was only the first pilot, on each landing, who had the heart rate increase. We recorded at least a 30-40% increase which was very stable and lasted for 30 or 60 s. The phenomenon does not qualify, by the way, as a startle response; if the landing has the characteristics of a startle you are in deep trouble.

WILSON: If I could just comment on data that I showed briefly from a larger study. During a formation landing, the pilots who flew in the lead position had heart rates which were higher than when they were flying in a wing position.

# EXPERTISE, STRESS, AND PILOT JUDGMENT

Christopher D. Wickens, Barbara Barnett, Alan Stokes, Tom Davis, Jr., and Fred Hyman
University of Illinois at Urbana-Champaign
Institute of Aviation, Aviation Research Laboratory
1 Airport Road
Savoy, IL 61874
USA

## SUMMARY

This paper describes two studies of pilot judgment examining the effects of stress and of expertise. Both studies were carried out on a computer-based aviation decision making simulation called MIDIS. In the first study the cognitive abilities of 40 instrument rated pilots, 20 novices and 20 experts were assessed. These pilots then flew the MIDIS simulator on a simulated cross-country flight during which their performance on a number of in-flight decisions was assessed. Experts were more confident than novices, but did not perform more optimally. The pattern of ability differences that predicted novice performance was different from that which predicted expert performance. In the second study, 10 instrument-rated pilots flew a different flight on MIDIS under conditions of stress (imposed by time pressure, noise, financial risk, and task loading), while 10 subjects flew in a nonstressed control condition. Stress had different effects on different kinds of decision problems. It degraded performance on those problems imposing high demand on working memory, but left unaffected those problems imposing high demand on the retrieval of facts from long term memory. The results are discussed in terms of the commonalities between the effects of expertise and stress, on the mechanisms of working memory and long term memory in pilot judgment.

## INTRODUCTION

Pilot judgment has been the subject of inquiry in a growing number of recent reports and papers, many of these published in the two most recent volumes of the Aviation Psychology Proceedings [12, 13]. Much of the growth in this research has been stimulated by Jensen's [11] seminal review article in which he pointed out that faulty pilot judgment has been a factor in a majority of aircraft accidents, a statistic that has not declined since the time of that report.

In considering the causes and limitations that lead to faulty pilot decisions, at least two would appear to be critical concerns: the experience level of the pilot and the conditions of stress in which the decision is carried out. Indeed it may well be that these two variables are linked in important ways. For example, it is possible that an important characteristic of expert decision making is that the expert is less susceptible to the degrading effects of stress. These two factors--experience and stress--will be the focus of the current investigations.

Many of the conclusions, assertions and statements regarding the causes of faulty pilot decision making are based upon anecdotal evidence and post-hoc accident analysis (e.g., [3], [15], [17], [18]). Indeed, post-hoc analysis has an important role to play (for example, in Air Force and NTSB accident reports), and because its data stem from the operational environment rather than from the laboratory, it is an important source of hypotheses. As a research method, however, this approach is less than fully satisfactory because such analyses are always subject to the 20-20 vision of hindsight, and often fail to consider the probabilistic nature of the environment in which any pilot operates. This probabilistic nature means that the "right" decision may occasionally produce an unfortunate outcome, and correspondingly, poor decisions may sometimes have no unfortunate consequences. Furthermore, it is well established that most accidents are determined by a cascading or confluence of multiple factors, only one of which may be related to a particular causal mechanism (e.g., stress, or lack of experience). Hence, it is hard to pinpoint these factors in isolation as being responsible agents.

It is indeed possible that plausible "intuitive" hunches regarding the effects of stressors and expertise on pilot decision making may be wrong. For example, with regard to expertise, a conclusion growing from recent work in decision making in other nonaviation environments is that increased practice on decision making tasks does not necessarily lead to better decision making, and may ironically lead to poorer decision making ([6], [1]). There is solid evidence as well that expert decision makers and forecasters are just as overconfident in the accuracy of their decisions as are novices, if not more so ([7]). To complicate matters further, the literature or stress contains ample evidence of tasks in which experimentally imposed stressors have left task performance unaffected, and perhaps even improved, depending upon the nature of the task ([2]).

The source of uncertainty in interpreting post-hoc analyses, and the ambiguity in interpreting stress and expertise effects results, in part, from the fact that decision making is a complex task, depending upon a number of separate cognitive abilities. Furthermore, the human's cognitive abilities that underlie decision making are themselves multifaceted. To help interpret this complexity, we describe below an information processing model of decision making which makes the important distinction between working memory (computational) and long term memory (direct retrieval) mechanisms in decision making. Decision problems may demand a mixture of the two components, and individuals, depending upon their level of skill, may utilize either or both mechanisms. As expertise increases, greater utility is placed upon direct retrieval. The model can help to interpret stress effects by detailing the information processing components upon which stress operates. Only decision problems that utilize stress-sensitive components will be expected to degrade. This paper is divided into four sections. In the first, we present the decision making model in greater detail. In the second, we describe a pilot decision making simulation known as MIDIS that has been used to test the model, and the effects of expertise and stress on decision making. The third and fourth sections describe our experimental results in examining these two influences, respectively.

## AN INFORMATION PROCESSING MODEL OF DECISION MAKING

Wickens and Flach ([19]) have proposed a model of pilot decision making that identifies different components in the decision process that may lead to errors or biases (see Figure 1). Prominent among these components are the use of selective attention to integrate cues relevant to the decision problem, the use of working memory, both verbal and spatial to entertain hypotheses and formulate spatial situation awareness, the use of long term memory from which to retrieve hypotheses, select choices, and evaluate risks, and a procedure to assess confidence.

A critical component of the model concerns its distinction in decision making between the role of working memory--the temporary, fragile, attention demanding system--and long term memory--the relatively permanent storage of knowledge. Working memory is employed in the "computational strategy" of decision making: entertaining and weighing hypotheses against the incoming cues, calculating the most likely hypothesis, and computing the choice of action with the highest expected utility. In contrast, using long term memory, a pattern of environmental cues may be directly matched with a stored, similar pattern, to reach a diagnosis through a direct "pattern match" ([16]). In turn, the recognized pattern may directly trigger the appropriate action through a stimulus-response association, a procedure which avoids or minimizes the fragile computational processes of hypothesis testing and choice optimization in working memory.

This shortcutting past the working memory system appears to have distinct implications involving both expertise and stress. In the first place it is reasonable to assume that as decision making expertise increases, pilots, having a large repertoire of past, similar experiences, will be more likely to avail themselves of the direct retrieval pattern match strategy, mapping the current set of symptoms onto a class of similar, previously experienced examples, a procedure noted by Ebbeson and Konecni ([5]) in judicial decision making, and by Rasmussen ([16]) in nuclear power plant failure diagnosis. This is not to say that the decisions will invariably be correct. Each decision problem is unique and the pattern of cues will not always make a perfect match to past experience, with the result that a misclassification may occur. But the choice would be made rapidly, with a relatively high degree of confidence.

Secondly, the difference between the fragile path through the computational procedures of working memory and the "direct retrieval" path through long term memory has implications for the effects of stress. Hockey ([9]) has identified a profile of stress effects on information processing associated with anxiety and high risk environments (i.e., characteristics of the pilot operating in dangerous conditions). This profile defines an increase in perceptual selectivity, a reduction in the capacity of working memory, a shift to more error prone responding, and a general increase in arousal, as consequences of anxiety imposed stress. Hence, the computationally-based decisions, depending on that working memory system, can be anticipated to degrade under stress. In contrast, we may predict that stress will leave relatively unaffected those of decisions that are made via the more "automatic" direct retrieval from long term memory.

## THE MIDIS DECISION SIMULATOR

Problem scenarios. Our vehicle for experimentally addressing the joint implications of stress and expertise on pilot judgment through the integrating framework of the model in Figure 1 is a microcomputer-based simulation of pilot decision making known as MIDIS. The MIDIS system is a microcomputer-based decision simulator implemented on an IBM PC/AT. MIDIS has a full, high-fidelity instrument panel based on that of a Beech Sport 180, the type of aircraft used for training at the University of Illinois Institute of Aviation. This display, implemented via the HALO graphics package and color enhanced graphics adapter, represents a full IFR "blind flying" panel with operating attitude, navigational and engine instruments. To enhance experimental validity, MIDIS has a number of simulator-like qualities (it provides a continuous "engine" sound cue, for example, and permits route deviations or reversals).

A team of flight instructors collaborating closely with cognitive psychologists has designed a series of flight decision problems or "scenarios" that incorporate the heterogeneous set of information processing demands that may be imposed upon the pilot. Generation of these scenarios has depended both upon an understanding of the pilot judgment model, and years of expertise in instrument flying. The scenarios are varied in terms of their qualitative and quantitative demand. For example, certain decision problems may require a breadth of attention, others may require that hypotheses be revised in light of new data, and still others may require an accurate assessment of risk. While incorporating these attributes, an effort has also been made to present the series of decision-situations as discrete events in a single coherently flowing flight from a real geographical origin to a destination.

The general structure of the MIDIS system places it in a class of programs referred to as "Graph Traversers" ([4]). Graph traversers are applicable to situations where a number of states are connected by a set of transformations or "operators." This can be represented as a branching tree-structure graph in which the nodes represent the states and the operators linking them are transitional probabilities. The states in MIDIS take the form of descriptions of realistic in-flight situations referred to as "scenarios." A scenario may involve any potential in-flight situation, emergency or otherwise. Each scenario requires that a decision be made among several alternatives presented. The decision influences the occurrence of subsequent scenarios since it selects the transitional probabilities that will operate. Therefore, as in the real world, poor decisions will typically lead to future choices in a less than ideal circumstance.

The scenarios presented to the pilots in our experiments were of two distinct forms, "static" and "dynamic," which were juxtaposed to create a continuous, coherent, and somewhat time-compressed flight. Static scenarios were presented via text above a static instrument panel configured for the appropriate stage of flight (e.g., level cruise) and the situation described (e.g., nearing a navigational fix). Dynamic scenarios presented a dynamically changing instrument panel in which the

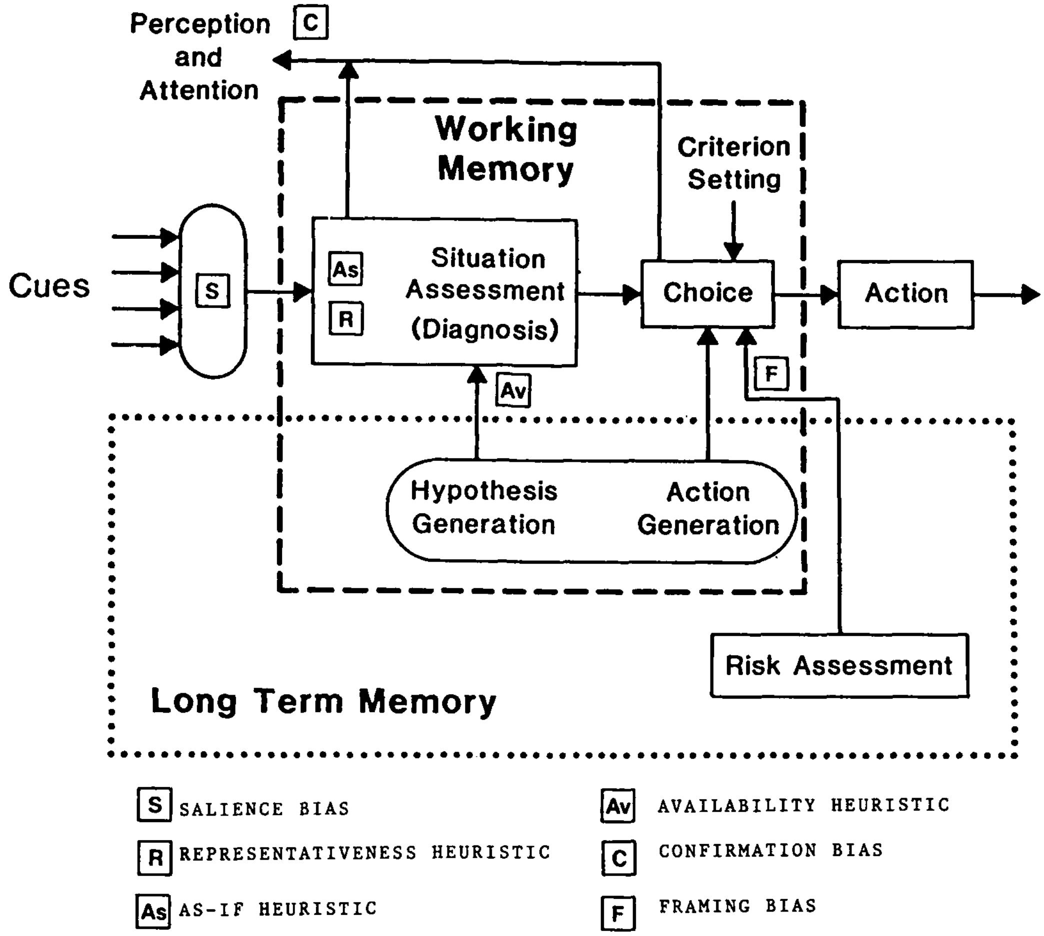



Figure 1: An information processing model of decision making. The boxes within the model indicate different heuristics and biases, whose identities are shown below (from [19]).

pattern of change sometimes indicated a potential problem or system malfunction. MIDIS problems encompassed the range of decision involving direct diagnosis of the different forms of instrument failures to judgmental calls of whether, for example, to continue with a landing or abort. Following the presentation of each problem, pilots are presented with a set of 4-5 options and must indicate by pressing the appropriate key, which option they believe represents the most optimal decision. This response is followed by a keyboard entry of confidence on a 1-5 scale.

Attribute and option coding. After finishing construction of the MIDIS scenarios, the flight instructors proceeded to generate two kinds of codes. First, each option in a decision scenario was assigned an optimality rating, on a scale from 5 to 1, in which the best option was assigned a value of 5. The less optimal options were assigned values ranging from 1 to 4, depending on how close they were to plausible alternatives. Second, the correct option in each scenario was assigned an attribute value code for each of the 10 critical cognitive attributes listed in Table 1. These attributes were selected based upon our content analysis of the flight scenarios in MIDIS, and guided by our expert analysis of pilot judgment. A value of zero indicated that the attribute was not relevant to the decision. Values from 1-3 indicated how critical it was for the subject to possess strength in the attribute in question, in order to choose the optimum option.

Cognitive battery development. One objective of the cognitive psychologists involved with the project was to develop a battery of cognitive tests that would match, as closely as possible, the attributes that were identified in each of the flight scenarios. Our efforts to identify existing cognitive tests that assessed these attributes, parallel an analogous effort performed by Irizarry and Knapp ([10]) in their study of individual differences in Army intelligence analysts. Based in part upon their study, and upon our own review of the literature, development of the cognitive battery proceeded in two phases. Initially, we compiled existing standardized tests (ETS Kit of Factor-Referenced Cognitive Tests, Eysenck Personality Inventory Items, MFF Test) that reflected measures on each of the relevant attributes. For those scenario attributes for which no standardized measure was readily available, specific tests were developed within our laboratory.

TABLE 1: Scenario Demands of Cognitive Attributes

1. Flexibility of Closure - the ability to find a given configuration in a distracting perceptual field.
2. Simultaneous Mental Integrative Processes - the ability to keep in mind simultaneously or to combine several premises or rules in order to produce a correct response.
3. Simultaneous Visual Integrative Processes - the ability to sample a select number of items from a complex visual display, and to combine this information in order to produce a correct response.
4. Sequential Memory Span - the ability to recall a number of distinct, sequential items from working memory.
5. Arithmetic Load - the ability to perform basic arithmetic operations with speed and accuracy.
6. Logical Reasoning - the ability to reason from premise to conclusion, or to evaluate the correctness of a conclusion.
7. Visualization of Position - the ability to perceive or maintain orientation with respect to objects in space, and to manipulate this image into other arrangements.
8. Risk Assessment and Risk Utilization - the ability to accurately assess the probability or riskiness of a situation, and to utilize this assessment in effectively carrying out decisions.
9. Impulsivity-Reflectivity - a measure of cognitive style differentiating those who tend to be fast and inaccurate (impulsive) or slow and accurate (reflective).
10. Declarative Knowledge - the ability to answer correctly a number of "textbook" questions covering a broad range of general aviation issues. This measure specifically excludes procedural or experience-based issues, focusing only on declarative facts and guidelines.

---

Thus, the compiled test battery consisted of a one-to-one mapping between cognitive attributes relevant to pilot judgment and cognitive tests specifically designed to measure each individual attribute. Table 2 provides a list of specific tests in the cognitive battery, as well as a description of the corresponding attributes. A more detailed description of the scenario attributes as well as specific tests in the cognitive battery may be found in Wickens et al. ([21]).

---

TABLE 2: Cognitive Test Battery

1. Hidden Figures Test (flexibility of closure).
2. Following Directions Test (simultaneous mental integrative processes).
3. Cue Sampling - Visual Integration Test (simultaneous visual integrative processes).
4. Visual Number Span Test (sequential memory span).
5. Subtraction and Multiplication Test (arithmetic load).
6. Nonsense Syllogisms Test (logical reasoning).
7. Surface Development Test, Card Rotations Test (visualization of position).
8. Risk Assessment and Utilization Test (need for *risk assessment and utilization*).
9. MFF Test and Impulsivity Self-Report Inventory (impulsivity and reflectivity - matched with latency and confidence).
10. Aviation General Knowledge Test (declarative knowledge).

---

## EXPERIMENT 1: INDIVIDUAL DIFFERENCES AND EXPERTISE IN PILOT JUDGMENT

In order to examine the effects of individual differences in cognitive performance and level of expertise on MIDIS performance, our first study tested a group of 40 instrument-rated pilots of varying degrees of experience.

### Method

The subject pool consisted of flight instructors from the University of Illinois Institute of Aviation, experienced instrument/commercial pilots with diverse backgrounds (e.g., Air National Guard, professional airline and private business flying), and Instrument Rated student pilots from both the Institute and local flight schools. The experiment is based upon a sample of 40 subjects divided into two cohorts, 20 pilots from the experienced group (greater than 400 flight hours) and 20 from the student group.

Data collection was conducted in two sessions for each subject. In the first session, lasting approximately 2 hours, the battery of psychological tests was administered. The second session involved the MIDIS simulation itself. Subjects were instructed to plan an IFR flight from Mountain View, Missouri, to St. Louis Regional Airport, planning the flight in their customary fashion with due regard for both the safety and efficiency of the trip.

Following the flight planning, pilots undertook the simulated MIDIS flight on the IBM PC/AT system. An initial practice flight trained subjects in the use of the color-coded keyboard and MIDIS conventions. The practice flight was not time-limited and could be re-entered and repeated until the subject felt comfortable with the system. After the practice flight and a reminder weather briefing, the flight from Mountain View to St. Louis was started. Each pilot's "flight" was unique in that the final destination and the route taken were dependent upon that pilot's decisions made along the course of the flight.

### Results and Discussion: Experiment 1

The analysis of decision performance in this study produced a number of interesting conclusions. First, in those elements of the decision making process tested by MIDIS, the expert pilots did not respond more optimally than the novices, although subjects in the former group were significantly more confident in their choices, as might be predicted by assuming that this group made greater use of direct long term memory retrieval strategies.

A second important finding related to the confidence ratings, is that both groups tend become less confident as problems became more difficult (i.e., on those problems in which their answers were less likely to be correct). However, the reduction in confidence with increasing problem difficulty was far less than optimal, thereby reflecting a classic pattern of overconfidence seen in many other domains of decision making (e.g., [7]). The degree of overconfidence was greater for the experts than for the novices.

A third finding was that the cognitive variables that predicted better performance for experts were different from those variables that predicted better performance for novices. In particular, while performance on dynamic problems was predicted for both groups by tests of the working memory capacity, substantial differences between the groups was found for the prediction of static problems. Variance in the performance of novices was related to declarative knowledge. But most variance in the performance of the experts was simply unrelated to any of the cognitive tests employed in the battery. This included tests of memory, attention and cognitive ability, as well as tests of declarative knowledge stored in long term memory (i.e., facts about aviation assessed through FAA questions and knowledge of aviation risks). We concluded that expert pilot judgment may be more heavily related to procedural knowledge or to pattern-recognition using the direct memory retrieval processes, than to the computationally intensive algorithms that would be predicted by tests of logical reasoning, memory and attention capacity. If in fact this is the case, then in accordance with the decision model in Figure 1, it may well be that certain aspects of pilot judgment are indeed relatively immune to stress effects, particularly for the expert pilot. It is this issue we address in the next experiment.

Finally, two other aspects of the predictive data from Experiment 1 are of note. First, our tests of risk assessment and utilization could reveal a differentiation between those who underestimated and those who overestimated risks. The latter group tended to perform more optimally on MIDIS. Secondly, the matching familiar figures (MFF) test provides an index of impulsive vs. reflective cognitive styles, and we observed that the more impulsive subjects on the MFF test also responded more rapidly on MIDIS (although without sacrificing accuracy).

In conclusion, the results of the first experiment suggest a distinction between those individuals, of lesser experience, whose decision making performance depends more upon the fluid capacity of their spatial abilities and working memory system, and those individuals with more extensive flight experience, whose decision performance was not well predicted by test battery measures. This dichotomy led us to infer that expert decision making is better predicted by the level of procedural knowledge, built up from experience in a way that allows direct retrieval of solutions from long term memory. It should be noted that this procedural knowledge is quite distinct from the declarative knowledge of instrument flying, assessed by the FAA written examination. This particular hypothesis is currently being examined in an ongoing experiment. However, by suggesting the importance of the working memory-long term memory dichotomy as a between-subjects variable, the results of Experiment 1 suggest that this same dichotomy could be important as a between problem variable in accounting for stress effects. This possibility was examined in the second experiment.

## EXPERIMENT 2: EFFECTS OF STRESS ON PILOT JUDGMENT

In the second experiment, two different groups of subjects carried out the MIDIS scenario, one in a nonstressed condition and the second while subjected to manipulations of stress.

### Method

Subjects in the stress conditions performed the MIDIS task under four simultaneous stress manipulations: (1) time pressure. Subjects were instructed to finish the flight in 1 hour (the mean time taken by the nonstressed group); (2) financial risk. Subjects were told that their earnings would be depleted for every minute in excess of this time; (3) dual task loading. Subjects were required to perform a secondary task, responding to visually presented letters on the instrument panel which were stimuli for a Sternberg memory search task ([20]); (4) noise stress. Failure to perform the loading task rapidly and accurately, produced an annoying 70 dB tone. However, even with successful completion of the secondary task, the tone was periodically (and randomly) presented. It was believed that the first two variables would induce a mild feeling of anxiety--the impending possibility of financial loss--while the second two would provide specific disruption of the working memory system.

The general procedures followed in Experiment 2 were identical to those in Experiment 1, and utilized a subset (20) of the same subject population that participated in Experiment 1. A different flight was programmed in MIDIS, from Saranac, New York to Boston's Logan Airport. The subjects in Experiment 2 had a mean experience level of 306 total flight hours and were subdivided into the two experimental groups. Effort was made to "match" pairs of subjects assigned to each group according to their level of flight experience and the similarity of their profile on the cognitive battery, which had been administered as part of their participation in the previous MIDIS study.

### Results and Discussion: Experiment 2

Analyses of the data revealed a significant degrading effect of the stress manipulation on decision optimality ($F_{1,9} = 6.41$; $p = 0.03$). Stress also reduced the mean confidence assigned to the response ($F_{1,9} = 5.18$; $p = 0.05$), but had no effect on the latency of performance.

The factor analysis of cognitive abilities from the earlier MIDIS study ([21]) had revealed three important "clusters" of the different cognitive attributes. These clusters were related to spatial demands, working memory demands, and knowledge demands. Our objective in the current research was to identify problems that were rated high and low on each of these attribute clusters. To assess spatial demands, the coded value of attributes related to flexibility of closure and visualization of position were summed for each scenario, and the scenarios were then assigned to one of two categories of spatial demand. The category values depended upon whether the sum was equal to 0 or 1 (low demand), or 4 to 6 (high demand). This categorization scheme assigned roughly thirteen scenarios each to the low and high spatial demand category.

A similar procedure was employed to categorize problems into two levels of dependency on stored knowledge (based on declarative knowledge and risk utilization), and two levels of working memory demand (based on attributes of simultaneous mental processes, sequential memory span and logical reasoning).

Table 3 presents the mean optimality scores for the two groups of subjects across the two demand levels, when those levels were coded by spatial demand, knowledge demand and working memory demand, respectively.

TABLE 3
Optimality Scores

| Attribute | Group | Attribute Demand | |
|---|---|---|---|
| | | Low | High |
| (Spatial) | | | |
| | Control | 3.35 | 3.30 |
| | Stress | 3.40 | 2.60 |
| (Knowledge) | | | |
| | Control | 4.30 | 3.25 |
| | Stress | 4.00 | 2.80 |
| (Working Memory) | | | |
| | Control | 3.40 | 3.50 |
| | Stress | 3.10 | 3.15 |

As evident from the first two rows of Table 3, optimality of decision making decreased with increasing spatial demand ($F1,9 = 9.73$; $p = 0.002$). More importantly this effect was manifest primarily in performance of the stress group, as revealed by the significant group x demand interaction ($F1,9 = 5.64$; $p = 0.042$). Therefore the data suggest that problems with high demand for spatial operations in memory are particularly sensitive to the degrading influence of the stress manipulation.

The data in the second two rows of Table 3 indicate the sensitivity of optimality scores to the demands imposed for general aviation knowledge ($F1,9 = 24.4$; $p < 0.001$). Greater requirements for general knowledge produced less optimal decisions. However, this effect was of roughly equivalent magnitude for both groups of subjects as there was no interaction between group and demand level. Supporting the predictions of the model, problems that were particularly dependent upon the retrieval of this knowledge from long term memory were not sensitive to the degrading effects of stress. Finally the third two rows of Table 3 present somewhat of an anomaly. Decision performance was essentially unaffected by the demand for verbal working memory, and this insensitivity was shown equally for stressed and control subjects alike.

One final observation concerned the effect of problem demand on confidence. While confidence ratings were generally unaffected by problem difficulty, there was exception in the case of knowledge demands. Here confidence actually increased as know. demands increased, even as the degree of optimality, for both groups, declined. This pattern reveals the classic finding of "overconfidence" typical of the findings of much decision making research ([6]).

## GENERAL DISCUSSION

The two experiments reported above reveal a pattern of results that is both internally consistent, and is consistent with other empirical findings in research on stress and decision making. Experiment 1 revealed that there was a different pattern of abilities underlying decision performance of novices and experts, a differentiation that was consistent with the dichotomy in the model between computational and direct retrieval processes in decision making. The results of Experiment 2, demonstrated that this dichotomy was equally important to predicting the effects of the imposed stressors on MIDIS decision making. Problems imposing a high demand on knowledge--therefore presumably employing direct retrieval for their solution--were not degraded by stress, whereas those imposing a high demand for computational processes in spatial working memory were so affected. Does this mean in turn that experts, more dependent upon direct retrieval for decision success, are less affected by stress? This issue is currently being examined in another experiment in our laboratory.

The internal consistency of these results must be qualified by two additional observations. First, we note that the decision attribute of declarative knowledge, the measure of direct memory retrieval that was found to account for stress resistance in the decision making of Experiment 2 was not the same as the procedural knowledge variable that was inferred to underlie expertise in Experiment 1. Specific tests of procedural knowledge are currently being developed in our laboratory. Secondly, the interpretation of the results of Experiment 2 was clouded by the failure to find that problems coded high on verbal working memory (as contrasted to spatial memory) were degraded by stress. It is possible here that our own coding of working memory demands was itself faulty since, unlike the other two attribute clusters, we failed to find any effects of working memory demand on decision optimality for either group.

Finally, the results of the two experiments were bound together by the effects of the manipulations on problem confidence. Confidence is a critical variable in pilot judgment, because the failure to appropriately calibrate, and the overconfidence that results, may lead the pilot to take unnecessary risks, or to fail to cross check the consequences of actions selected. In the current research we found that experts tended to be more confident than novices, to the point of overconfidence, and correspondingly that problems that placed high demands on direct memory retrieval tended to be answered with more confidence than those that did not (even though the former were in fact responded to less optimally than the latter). We found that increasing levels of stress reduced confidence.

In summary, the data reported here are consistent with the information processing model of decision making presented in Figure 1. In its current form, the model falls far short of the useful computational models in such areas as multi-axis manual control ([14]), and/or visual target acquisition ([8]). Indeed, given the complexity of decision making, it is unlikely that computational models will ever be well-developed in this area. Nevertheless, the important feature of the model is that it can provide an organizing framework for understanding the influences on pilot judgment caused by such things as the kind of problem, the skill level of the pilot, and the environmental influences leading to the perception of stress.

## REFERENCES


1. Brehmer, B., Models of diagnostic judgment. In J. Rasmussen & W. Rouse (Eds.), Human detection and diagnosis of system failures. New York: Plenum Press, 1981.

2. Broadbent, D., Decision and stress. London: Academic Press, 1971.

3. Connolly, T.J., Blackwell, B.B., & Lester, L.F., A simulator-based approach to training in aeronautical decision making. Proc., 4th Int'l Symp. on Avia. Psych., Ohio State University, Dept. of Avia., Columbus, OH, 1987.

4. Doran, J.E., & Mitchie, D., Experiments with the graph traverser program. Proc. R. Soc. (A) 294, 1966, 235-259.

5. Ebbeson, E.D., & Konecni, V., On external validity in decision-making research, in T. Wallsten (Ed.), Cognitive processes in choice and decision making. Hilldale, NJ: Erlbaum Associates, 1980.

6. Einhorn, H., & Hogarth, R., Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 85, 1978, 395-416.

7. Fischoff, B., & MacGregor, D., Subjective confidence in forecasts, Woodland Hills, CA: Perceptronics (Technical Report PTR-1092-81-12), 1981.

8. Greening, C.P., Mathematical modeling of air-to-ground target acquisition. Human Factors, 1976, 18(2), 111.

9. Hockey, G.J.R., Changes in operator efficiency. In K. Boff, L. Kaufman, & J. Thomas (Eds.), Handbook of perception & performance (Vol. II). New York: John Wiley, 1986.

10. Irizarry, V., & Knapp, B., A preliminary investigation of problem-solving and judgmental strategies of expert military intelligence personnel. Proc., Psychology in the Dept. of Defense. Colorado Springs, CO: USAFA, 1986.

11. Jensen, R.S., Prediction and quickening in prospective flight displays for curved landing and approaches. Human Factors, 23, 1981, 333-364.

12. Jensen, R.S., & Andrion, J. (Eds.), Proc., 3rd Int'l. Symp. on Avia. Psych., Ohio State University, Dept. of Avia., Columbus, OH, 1985.

13. Jensen, R.S., (Ed.), Proc., 4th Int'l. Symp. on Avia. Psych., Ohio State University, Dept. of Avia., Columbus, OH, 1987.

14. Levison, W. H., Baron, S., & Kleinman, D.L., A model for human controller remnant. IEEE Trans. in Man-Machine Systems, MMS-10, 1969, 101-108,

15. Lubner, M.E., & Lester, L.F., A program to identify and treat 'pilot error,' particularly, poor pilot judgment. Proc., 4th Int'l. Symp. on Avia. Psych., Ohio State University, Dept. of Avia., Columbus, OH, 1987.

16. Rasmussen, J., Models of mental strategies in process control. In J. Rasmussen & W. Rouse (Eds)., Human detection and diagnosis of system failures. New York: Plenum Press, 1981.

17. Simmel, E.C., Cerkovnik, M., & McCarthy, J.E., Sources of stress affecting pilot judgment. Proc., 4th Int'l. Symp. on Avia. Psych., Ohio State University, Dept. of Avia., Columbus, OH, 1987.

18. Simmel, E.C., & Shelton, R., The assessment of nonroutine situations by pilots: A two-part process. Avia., Space, & Environ. Med., 1987, 1119-1121.

19. Wickens, C.D., & Flach, J., Human information processing. In E. Wiener & D. Nagel (Eds.), Human factors in aviation. New York: Wiley, 1988.

20. Wickens, C.D., Hyman, F., Dellinger, J., Taylor, H., & Meador, M., The Sternberg memory search task as an index of pilot workload. Ergonomics, 1986, 11, 1371-1383.

21. Wickens, C.D., Stokes, A., Barnett, B., & Davis, T., Jr., University of Illinois, Institute of Avia., Avia. Res. Lab., Componential analysis of pilot decision making, 1987, Final Technical Report (ARL-87-4/SCEEE-87-1).

ACKNOWLEDGMENTS

This research was supported by subcontract No. DOE EGG C87-101376-2 from Idaho National Engineering Lab as part of a larger effort funded by the Air Force Aeromedical Research Lab Human Engineering Division. Ms. Heidi Hahn and Mr. Gary Reid are the technical monitors of the two organizations, respectively. The authors wish to acknowledge the programming contributions of Tak-Ming Lo and Rob Rosenblum of the MIDIS simulation.

## DISCUSSION

URSIN: Is there any chance that the split that you had on your last overhead is related in any way to hemispherical differences? Could you account for it by spatial versus digital type of differences in the memory capacity?

STOKES: Well, we were concerned because our secondary loading task was a visual task. Another interpretation of the results would be that the decrement and decision optimality came largely from decrements on the dynamic scenarios rather than the static ones. The subjects were monitoring a rate of change in the instruments, trying to detect a fault and timesharing with some text giving a heading and the phase of flight. At the same time they were trying to detect a Sternberg prompt on the screen. So in our follow on study we are looking at ways of presenting the secondary task through a different channel by using an auditory Sternberg or some comparable task.

# GENERAL DISCUSSION

STEVENSON: I would like to make a comment about a previous paper because I did not think about it earlier. We were told this morning by Wing Cdr Spiller that we need pilots who are aggressive, self reliant, do not express their emotions well and prefer action rather than contemplation. Yet perhaps the ideal fighter pilot of the future will be a chess master and a computer game strategist. We are also told that massive inputs of information are going to load even the traditional fighter pilot. The obvious conclusion is to put the first kind of man in the front cockpit to fly the aeroplane, fly the dog fight, avoid the ground, and put the second kind of man in the back seat to work the computers and engage the enemy at long range. One of the problems is that the first kind of pilot tends to distrust and dislike the kind of guy that you have put in the back seat. They tend to trust their wing-men because they are also fighter pilots. They are a known factor and they are like themselves; they are somebody they can trust, but they do not necessarily trust the guy in the back seat.

The talk we had this afternoon by Dr Foushee about airline pilots, the aggressive, 'right stuff' fellow who didn't do too well at the beginning but whose ability and his quick response to emergencies increased later. Perhaps this suggests that the fellow does not trust the other person's ability and wants to rely on his own; but in time he does learn that he can trust the guy in the back seat. I wonder if this human factor problem will become more important in future two seat aircraft. Aggressive, independent guys in the front seat are going to have to learn to deal with people in the back seat who, basically, will rule their lives and assure their survival, or lack of it. We are already having trouble in the United States Air Force with the F-16s and the F-4s that are working together in a 'Wild Weasel' role. The F-16 fighter pilots resent greatly having to function as an accessory to the F-4 Wild Weasel and take direction from the Wild Weasel Wizzo in attacking their targets. They really do not like this. I just wonder if this type of conflict is one that we are going to run into more and more in the future.

BILLINGS: I remember at an earlier AGARD meeting we had the privilege of going to the NBB at München where we were briefed by the Chief Test Engineer and the Chief Test Pilot for Tornado. One comment that particularly struck me was that the Chief Test Pilot made some comment that the aircraft was a wonderful joy to fly. The only problem was he never had any idea where the hell he was. Whereupon the Chief Engineer got up and said "Well that's all right because I never have any idea what he is about to do either", speaking about the guy in the front seat. They were sort of dependent on each other in that both were going to the same place at more or less the same time. They also pointed out that the terrain following radar had some really marvellous capabilities, that it could go to very low altitude at very high speed. At which point the Test Pilot remarked that, of course, it had not been flown at that limiting altitude and that limiting speed yet. He thought perhaps he would try it sometime. Whereupon the Chief Test Engineer in the back said: "If he is crazy enough to do it I am not going to be there". So I think your idea of a certain degree of mistrust is probably correct, but at the same time, I think we are going to be totally dependent on a second human wherever he happens to be located. It is really high time that we looked into the implications of such a relationship.

SPILLER: I started proceedings so I thought I ought to just chip back here. I speak as a man with 2000 hrs single seat and 2000 hrs F-4 so I have a balanced view on the world. A good navigator is a great help. A bad one is a pain in the arse; he is absolutely no use at all. I am certain that you are right. Two people are needed in an aeroplane of the future, but Europe is certainly destined to having a single seat aeroplane certainly to the year 2020. I think we need to bear that in mind.

# STRESS AND PERFORMANCE DURING A SIMULATED FLIGHT IN A F-16 SIMULATOR

Ragnar J. Værnes, Marit Warncke, Grete Myhre and Asbjørn Aakvaag
The Norwegian Underwater Technology Centre (NUTEC),
P.O. Box 6, N-5034 Ytre Laksevåg,
The Hormone Laboratory, Haukeland Hospital, RNoAF, and
Institute of Aviation Medicine, Oslo, Norway

## SUMMARY

Sixteen Norwegian F-16 pilots (average age = 24.2 years, average experience = 2.2 years) were tested before, during and after a 90 min flight in a F-16 simulator. During the flight different emergency operations and landings in difficult weather conditions had to be performed. The pilot performance was logged continuously during the flight (Accepted/Not accepted). Heartrate (HR)/heartrate variability (HRV) was monitored continuously. Saliva for cortisol analysis and urine for catecholamine analysis were sampled before and after the flight. Tests of anxiety (state and trait) and defense mechanisms (Plutchik's "Life Style Index", LSI) were administered. In addition to the pilots' testresults on Krag's "Defense Mechanism Test" (DMT), results on psychomotor performance from the selection-period were used. The endocrine and the HR-results indicated that the pilots were very activated during the flight. A HR of 120 beats/min was registrated. There were significant correlations between endocrine levels and not-accepted performance. Pilots with high defense mechanisms were significantly less activated on HRV, but had more pilot errors. Pilots with high defense considered simulator training as less important and they also trusted more the instruments in the aircraft. There was a significant correlation between high defense and number of near miss episodes during real flights. The results confirm previous studies which have shown that high defense correlates both to endocrine activation and impaired performance during stress in high risk occupations.

## INTRODUCTION

Data have been published periodically since 1940 showing that three out of four aircraft accidents apparently result from inadequate performance of the human component in the aircraft man-machine system (1, 2). This proportion has persisted in spite of years of exhortation to pilots to perform more consistently and with fewer errors. The exhortation has been notably, but predictably, ineffective in modifying the relative magnitude of this source of accidents (2).

The errors made by pilots are in principle no different from those made by everyone else. The use of the word "pilot error" has suggested that somehow the nature of the errors made by this kind of operator is unique; that once an accident could be attributed to this "cause", then the problem was solved and the case could be filed. Today the expression "human factor" is preferred.

The concept of accident proneness originated early in the 20th century (3). Accident proneness is the tendency of some people to have more accidents than others with equivalent risk exposure, for reasons beyond chance alone. It is clear that one person can have more accidents than another purely by chance. One other reason may simply be the one of exposure to risk. A pilot flying routinely in bad weather with poorly equipped aircraft and landing on inadequate airfields is exposed to more risk than one flying in nice weather with modern avionics.

In addition to pure chance and exposure, a third reason for variation in the individuals accident record could be that he possesses some innate characteristics (or "personality") which make him more liable to accidents. Some scientists see this as a more appropriate definition of accident proneness.

A complicated emergency situation in an aircraft can be met by direct action (coping) or lead to a defensive distortion of the perception of the situation (4). The defense-concept has existed in ego psychology since Freud, and has been used extensively as an explanatory concept in clinical work. Originally, the classical defenses provided an identification of ten processes, or mechanisms as they are called in psychoanalytic terms. Defense is most likely to occur when no means of coping is available (5), and the level of fear is high. The defensive distortion of the perception reduces efficiency in handling the dangerous situation (6).

The use of coping and defense concepts in stress research has led to a shift in emphasis from the amount of anxiety aroused or the intensity of the threat, to the various ways in which people handle the challenging (threatening) situation. The absolute intensity of arousal has often been the target of measurement, but this "perceptual ego-model" suggests that the same amounts of affect can occur for different reasons, and that it may have different consequences on how individuals handle this affect. There is considerable evidence for this position in empirical research. For a review see Værnes (7).

In this paper the authors report data on the relationship between personality factors such as defensive strategies, flight performance and psychophysiological reactions during simulated flights in a F16-simulator. The results are discussed in view of the theories on the relationships between personality factors and performance in hazardous situations.

# METHOD

## SUBJECTS

Sixteen male F-16 fighter pilots from the Royal Norwegian Airforce (RNoAF) participated. Average age was 24.2 years (1 sd = 1.3), they had been pilots for 3.6 years (1sd = 0.7) and had been flying jetfighter for 2.2 years (1 sd = 0.9).

The pilots had in average 76.6 hours on F-5 (1 sd = 53.1) and 197.8 hours on F16 (1 sd = 149.2). They had previously 51.3 hours on the F-16 simulator (1 sd = 23.3). In average it was 34 days since last session on the simulator (1 sd = 34.1).

None of the 16 pilots had previously near-misses with loss of aircraft. However, some had experienced episodes such as flame out, near-collision, and puncture during take-off/landing.

## INSTRUMENTS

### The Flight Simulator

F-16A is a single-engine, single seat, multirole tactical fighter with full air-to air and air-to-surface combat capabilities. The fire control system includes a fire control radar with search and tracking capability, a radar electrooptical (REO) display, and a head-up display (HUD). A stores management system (SMS) presents a control panel and visual display for inventory, control, and release of all stores. The cockpit is conventional except for the seat, which is reclined 30 degrees, and the stick, which is mounted on the right console.

The flight simulator and the operator console were in separate rooms. The pilot used his private helmet, mask and G-suit during the simulated flight. He had radio communication with the system operator and a pilot who functioned as flight - controllers after standard operational procedures.

### Psychological Tests

Life Style index (LSI, 8): Defense strategies were measured by a Norwegian translation of LSI. The test consists of 92 items in the form of 'true' or 'false' statements. Scores can be calculated for each of eight subscales (denial, repression, regression, compensation, projection, displacement, intellect-ualization, and reaction formation). High scores indicate strong defense strategies. In addition, an overall score can be calculated (LSI SUM).

Defense Mechanism Test (DMT, 9): The DMT is one of the psychological selection instruments in the RNoAF. The test is administered after the screening from the different group-tests. Of the remaining 50-40 %, about 20% are screened out on DMT. Despite this screening , there is still a wide variation on the results for psychological defense mechanisms among the pilots.

The DMT consists of repeated exposures (20 in all) of "Thematic Apperception Test" - like pictures in a tachistoscope (TAT, 10). The exposure time is increased gradually by steps from 10 to 2000 ms. Two such series are given. The subject is not informed that the same picture is exposed repeatedly. After each exposure the subject makes a sketch of what he has seen, and gives a verbal report. The responses are classified on the basis of their thematic content and the development of the percepts in the course of the series of exposures. The scoring follows a procedure described in detail by Kragh (9, 11) and Værnes (12). Since the pilots previously were selected on Neumans Pilot Index (NPI) and 10-aspect, the variable DMT DEF were used here indicating the overall use of defensive strategies throughout the DMT-series.

State/Trait Anxiety (A-STATE/A-TRAIT, 13): A-State is conceptualized as a transitory emotional state or condition that is characterized by subjective, consciously perceived feelings of tension and apprehension, and heightened sympathetic nervous system activity. A-States may vary in intensity and fluctuate over time (13). The A-State scale consists of 20 statements, and the subjects are asked to indicate how they feel at the particular moment when filling in the questionnaire. Trait anxiety (A-Trait) refers to the relatively stable anxiety proneness; to the tendency to respond to situations perceived as threatening with elevations in A-State intensity. The A-Trait scale also consists of 20 statements but asks people to describe how they generally feel. For both scales, the total scores were used.

Intellectual Abilities and Psychomotoric Performance: In addition to the DMT from the pilot uptake, the results on general IQ and the average score on all the psychomotor tests were used. For both variables a 9-point scale was used with high score indicating good performance.

### Physiological Tests

Heartrate and Heartrate-variability (HR and HRV): HR was monitored using a portable Oxford Instrument recorder (Oxford Electronic Instruments, series 4.24). Electrodes (NIKO) were fastened on the lower rib on both sides, and the recorder was mounted on the flight-suit so it would not interfere with performance. Average HR was calculated for one minute periods at fixed schedules: When a new task was introduced during the flight (see Performance Measurements). The HRV was calculated in milliseconds (ms) as beat-to-beat signals for the same periods.

Endocrine Measures: Cortisol (nM) and Testosterone (pM) were measured in saliva. For Cortisol a standard kit from Farmos Diagnostics, Abo, Finnland was used. The analysis was run after recommended procedures for saliva. Testosterone was analysed using a kit from Radio Assay System Laboratories, Carlson, California. One ml of saliva was extracted with 6 ml of diethlether, the extract evaporated to dryness, and the residue reconstructed to zero standard. Adrenalin and Noradrenalin were analyzed in the urine using a fluormetric method (nanogram/minute).

**Performance Measures**

The flight lasted about 80 minutes in the simulator. All pilots went through the same tasks, and the performance criteria for each task were if he succeeded or not. The seven following tasks was administered throughout the flight:

1) Asyncrone flaps: This was introduced at stable altitude during the flight to Bardufoss. Not acceptable performance: Did not lock the flaps according to manual.

2) Landing on Bardufoss: Due to the mountains it is quite complicated to land at this airport. In addition the pilot had to land under minimal weather conditions. Not acceptable performance: Had to fly over or crashed.

3) Bombing: The RNoAF has an area for bombing training. The task was to drop 3 bombs within a radius of 50 metres. Not accepted performance: Failed to hit the target.

4) Intercept: An enemy-plane should be intercepted at 20000 ft and shot down. Not accepted performance: Did not manage to perform the task.

5) Landing on an extreme short runway: The pilot should land and stop completely. Not accepted performance: Did not manage to stop, and had to eject.

6) Flame-out: A flame-out at 8000 ft was introduced. The pilot had about 50 seconds to restart. Not accepted performance: Did not manage to start and had to shoot himself out.

7) 1. and 2 hydraulic failure: During landing at Bodø, the front-wheel got stuck. While the pilot searched the manual a hydraulic failure was introduced so all control over the aircraft was lost. Not accepted performance: Did not manage to eject in time.

**Subjective Evaluation**

The pilot was asked to evaluate each task after the flight on a 5-point scale from 'none' to 'very' regarding 1) Difficulty, 2) Realistic, 3) Satisfaction with the instrument-panel, and 4) Perceived stress.

EXPERIMENTAL PROCEDURES

Prior to the data collection, permissions from RNoAF and the Squadron commanders of the two squadrons involved were obtained. All pilots were informed about the project. This was necessary to motivate them to participate. Care was taken, however, not to bias questionnaire answers, and the 'stress' concept was treated in very general terms.

Confidentiality of data was secured by using code numbers on all data sheets and keeping the key to the code separately from the data.

Testing prior to flying: The pilot was transported to the simulator-building by car. Prior to the flight he was given a code number after signing a statement of willingness to participate in the project. This was followed by structured interview regarding previous experience as a pilot.

A-State was then filled in, and urine and saliva samples collected. Electrodes for HR/HRV was mounted and baseline measure was performed.

A standard briefing about the flight was performed with the pilot/operator.

Testing during the flight: After finishing all preflight procedures clearance was given for takeoff and heading Bardufoss. All communication was on standard radio-frequencies. In addition to a standard protocol which was filled in during the flight, pilot-errors could later be analysed from computer memory.

Testing after the flight: New samples of urine and saliva were immediately collected. The questionnaire regarding Subjective Evaluation, the LSI and A-Trait were then administered.

STATISTICS

Data were analysed with SPSS/PC + (14, 15). Standard correlations, t-testing and frequency-analysis were performed.

**RESULTS**

THE DESCRIPTIVE ANALYSIS

**Psychological Tests**

Life Style Index (LSI): The mean total score on the defense scale (LSI SUM ) was 23.0 (1 sd = 8.8). This score is comparable with those reported by two previous 'high-stress group' studies on fighter pilots (16) and divers (17). These groups have a somewhat lower LSI TOT than in other 'ordinary occupations' such as shift workers (18) and nurses (19).

Anxiety: Mean A-Trait was 31.4 (1 sd = 5.1) which also compared favourably with the average scores for the previously reported studies, i.e., on fighter pilots (16). Average A-State prior to the flight in the simulator was 31.4 (1 sd = 5.1). This is somewhat higher than in the previous fighterpilot study (16) where flight-performance not was a part of the study.

**Physiological Tests**

Adrenalin and Noradrenalin: The average level of Adrenalin prior to the flight was 75.3 ng/min (1 sd = 62.2). Postflight, the average score was 91.4 ng/min (sd = 62.2). The standard deviation, especially for the preflight average indicated a marked individual activation in conjunction with the simulator flight.

The average level of Noradrenalin prior to the flight was 129.8 ng/min (1 sd = 130). Postflight, the average score was reduced to 104.1 ng/min (1 sd = 86.6). Prelevels of both Adrenalin and Noradrenalin were high compared with other studies.

Cortisol and Testosterone: The average Cortisol-level confirmed the results on the catecholamines: Preflight level was 15.2 nM (1 sd = 8.5) and postflight level was 17.2 nM (1 sd = 6.5).

The average level of Testosterone was 213.7 pM (1 sd = 67.4) preflight, and 219.9 pM (1 sd = 64.1) postflight.

Heartrate (HR) and Heartrate variability (HRV): Average HR was relatively high throughout the whole flight (see Figure 1).

After the crash the average HR was up to 103 beats/min. The stipled curve indicates that maximum HR was over 120 beats/min during phases of the flight (see Figure 1).

The activation led also to a relatively high HRV (see Figure 2). In conjunction with several of the tasks, the average HRV was over 50 msec and maximum HRVs were up to 150 msec, indicating a marked stress reaction in some pilots.

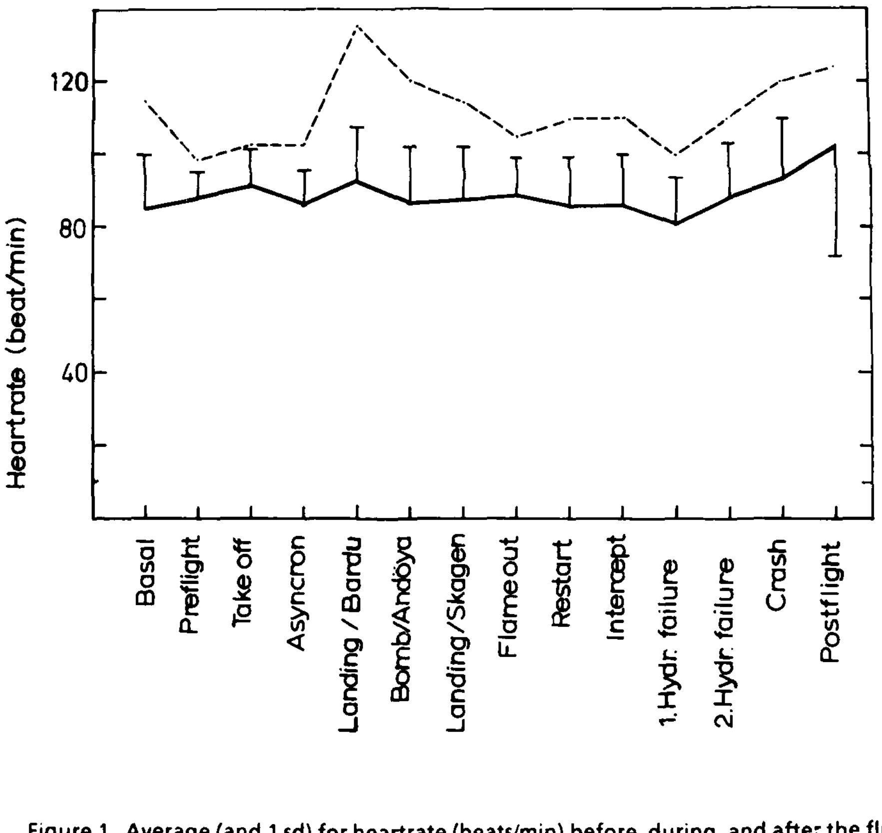



Figure 1. Average (and 1 sd) for heartrate (beats/min) before, during, and after the flight in the F16 simulator. The stipled curve indicates maximum heartrate.

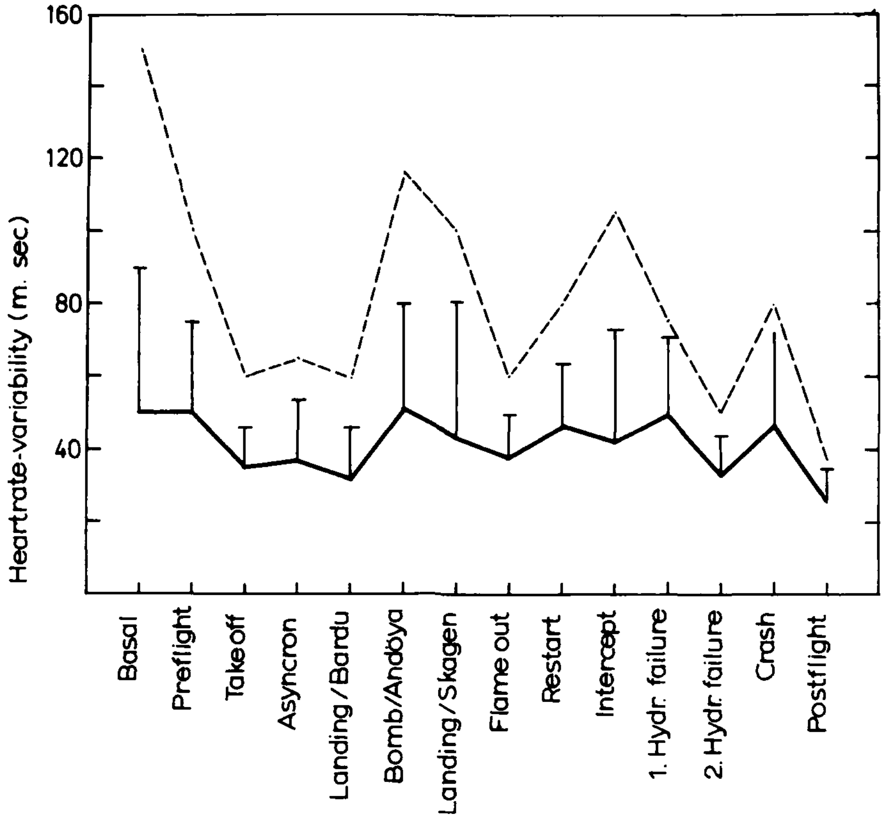



Figure 2. Average (and 1 sd) for heartrate-variability (millisec) before, during, and after the flight in the F16 simulator. The stipled curve indicates maximum heartrate variability.

**Performance Measures**

Number of pilots who succeeded on the different tasks throughout the flight are shown in Table 1. Since some of the pilots had to shoot himself out or crashed early in the flight there is a lower total sum than 16 on the last task.

Table 1. Number of pilots who succeeded/ did not succeed on the different tasks.

| Task | Succeeded | Did not succeed |
|---|---|---|
| Landing at Bardufoss | 9 | 7 |
| Bombing | 14 | 2 |
| Intercept | 11 | 5 |
| Landing on short runway | 14 | 2 |
| Restart after flame-out | 12 | 4 |
| Ejected due to 1. & 2. failure | 10 | 2 |

As many as 7 pilots did not manage to land at Bardufoss under minimal weather conditions. The bombing and landing on short runway were, however, easier. Nearly one third of the pilots did not manage to restart after flame-out and have a proper intercept of an enemy-plane. Two of the remaining 12 pilots did not eject early enough when the 1. and 2. hydraulic failures were introduced.

**Subjective Evaluation**

On the question: "To what extend to you feel the simulator training important/useful" the average score was 4.6 (1 sd = 0.5) on a 5 point scale from "none" to "very useful". The pilots indicated that the last task was the most complicated, followed by the landing tasks on short runway and landing at Bardufoss (see Table 2).

Table 2. Average (and 1 sd) score on difficulty, realism and satisfaction with instrument on the different tasks on a 5-point scale from none to very.

| Task | Difficulty | Realism | Satisf. with instr. |
|---|---|---|---|
| Landing at Bardufoss | 3.1 (1.0) | 3.3 (1.3) | 4.4 (0.5) |
| Bombing | 2.3 (0.7) | 2.2 (1.2) | 4.7 (0.5) |
| Intercept | 2.2 (1.3) | 3.6 (1.6) | 4.4 (0.7) |
| Land on short runway | 3.2 (1.1) | 3.5 (1.2) | 4.8 (0.5) |
| Restart after flameout | 2.6 (1.4) | 3.5 (1.6) | - |
| Ejection | 3.9 (1.4) | 2.4 (1.7) | 4.4 (0.9) |

The bombing and the introduction of 1. and 2. hydraulic failure were considered less realistic than the other four tasks. In general, the pilots were quite satisfied with the instrumentation on all tasks with a tendency that the instrument functioned somewhat better during short runway landing and bombing.

THE CORRELATION ANALYSIS

1. WHAT ARE THE RELATIONSHIPS BETWEEN THE PHYSIOLOGICAL ACTIVATION AND PERFORMANCE?

Several negative correlations between performance and endocrine activation were obtained. Especially pilots who failed on short runway landing and intercept had high preflight levels of cortisol (r = .43 to -.62, p < .05 to .01). The same pattern was found for Adrenalin (r = .75, p < .001) and Noradrenalin (r = .80, p < .001). The Noradrenalin-level correlated also to not-successful landing at Bardufoss (r = -.75, p < .001) and to not-successfull restart after flame out (r = -.75, p < .01).

Regarding the relationships between the heartrate-variables and performance, the following correlations emerged. On five of the tasks not-accepted performance correlated to HR (r = .43 to -.55, p < .05 to .01). The same was found for HRV (r = ÷ .45 to -.55, p < .05 to .01).

2. TO WHAT EXTENT DID EXPERIENCE PLAY A ROLE ON FLIGH-PERFORMANCE AND STRESS-REACTIONS

Pilots with less experience (number of years with pilot licence) did not manage to land at Bardufoss (r = -.43, p < .05). An interesting finding was that pilots with long interval since last simulator training did not manage the flameout task (r = -.62, p < .01), the short runway landing (r = -.43, p < .05) and the ejection (r = -.66, p < .01). Pilots who had a high score on the necessity of simulator training significantly succeeded more on the last task (r = .58, p < .01).

On the heartrate-variables an interesting pattern occurred: While there were several negative correlations between experience and HR at the beginning of the flight (r = -.46 to -.65, p < .05 to .01), the opposite was found during the last part of the flight (r = .58 to .66, p < .05 to .01). I.e., while the inexperineced pilots were most activated at the start, the experienced pilots got more activated during the more complicated/serious tasks such as flameout, interception and ejection.

This was confirmed by the correlations with the postflight endocrine levels: Number of hours on F-16 correlated significantly with the Cortisol-level (r = .67, p < .01). Noradrenalin postflight correlated significantly with number of days since last simulator session (r = .51, p < .05) and with number of real accidents/near-misses (r = .56, p < .05).

3. WHAT ARE THE RELATIONSHIPS BETWEEN THE PSYCHOLOGICAL FACTORS, PERFORMANCE AND PHYSIOLOGICAL ACTIVATION?

There were significant negative correlations between accepted intercept and LSI SUM (r = ÷ .66, p < .01) and between landing on short runway and LSI SUM (r = ÷ .52, p < .05). Pilots with high LSI defense underreacted also on HR during the flight. There were 15 significant negative correlations between defense and HR (r = -45 to -.71, p < .05 to < .01) while 5 could be expected by chance.

This was more prominent for HRV: While 5 significant correlations could be expected by chance, 26 significant correlations were obtained between HRV and LSI SUM (r = -.44 to -.70, p < .05 to < .01).

An interesting finding was that pilots with high LSI defense considered simulator training less necessary (r = -.65, p < .01), and therefore it was a longer time since they last trained (r = .46, p < .05). The pilots with high LSI defense trusted more/were more satisfied with the instruments: There were 11 significant correlations between these statements for the different tasks and high defense (r = .43

to .63, $p<.05$ to $<.01$) while 2 could be expected by chance. Pilots with high score on defense mechanisms considered also that the simulator flight to be less realistic and less complicated than the other pilots ($r = -.44$ to $-.,55, p<.05$).

4. HOW DID THE PILOT'S EVALUATION REGARDING DIFFICULTY AND PERCEIVED STRESS RELATE TO PERFORMANCE AND THE PHYSIOLOGICAL REACTIONS?

   Perceived stress in the simulator correlated to the preflight levels of Cortisol ($r = .48$, $p<.05$). Perceived realism and difficulty correlated negatively with accepted performance ($r = -.55$ to $-.83$, $p<.05$ to $.001$). This should indicate that lack of motivation was not a significant factor for poor performance during this test-flight.

   Twentyone significant correlations between HRV and perceived realism for the flameout-, intercept-, and Bardufoss-tasks were obtained while 3 could be expected by chance ($r = .45$ to $.72$, $p<.05$ to $.001$).

5. WHAT WERE THE RELATIONSHIPS BETWEEN THE RESULTS ON THE SELECTION VARIABLES ON PILOT-SELECTION AND THESE RESULTS ON THE SIMULATOR?

   Several interesting relationships to high DMT defense were obtained: Pilots with high DMT DEF were more satisfied/trusted the instrument panel more during the different tasks ($r = .52$ to $.54$, $p<.05$). These pilots, especially those with high scores on Reaction Formation, reported also that the different emergency tasks were less difficult ($r = -.45$ to $-.46$, $p<.05$). The general perceived stress was also negatively correlated with DMT DEF ($r = .-51$, $p<.05$).

   The finding that pilots with high defense scores (both LSI and DMT) trusted the instrumentation more and (therefore) reported that the different tasks were less complicated are in contrast to the finding that high defense (DMT) was significantly correlated to number of real accidents/near misses ($r = .59$, $p<.05$).

   Even if DMT DEF was negatively correlated with NPI and 10-aspect, did these newer constructed variables not relate to the simulator-performance. The only findings indicated that good NPI and 10-aspect correlated with postflight levels of Cortisol ($r = .53$, $p<.05$) and with perceived difficulty during the flight ($r = .55$ to $.66$, $p<.01$).

   General IQ and results on the psychomotoric tests during uptake were positively correlated to perceived difficulty ($r = .54$, $p<0.5$), with HR ($r = .44$, $p<05$) and HRV ($r = .44$, $p<05$) during the last task when they totally lost control over the aircraft.

## DISCUSSION

These pilots had an average defense and anxiety score lower than people in other occupations tested. They were at the same level as for other jetfighters and crew on Hercules (16) and for deep sea divers (17). The endocrine and heartrate variables indicated, however, that the pilots were highly activated immediately prior to and during this testflight. Some of them had HR of more than 120 beats/min during some of the tasks.

Flameout, intercept and landing at Bardufoss were more complicated than the other tasks, and 25% to 45% of the pilots did not manage successfully here. The subjective evaluation of the different tasks indicated that especially the flameout, intercept and the short runway landing were realistic.Even if the pilots wre generally satisfied with the instrumentaiton, this was somewhat higher scored for bombing and short runway landing.

Significant correlations were obtained between not-accepted performance and high endocrine level, especially for intercept and landing at Bardufoss. The same pattern was found for HRV. Quite naturally, experience in the F16 correlated postively to performance, such as on the flameout task and the quick ejection after 1. and 2. hydraulic failure. The positive correlations between physiological activation and experience during the last, and more "dramatic" tasks indicated that the experienced pilots got more emotionally aroused. This was also confirmed in the postflight endocrine measures.

Regarding the personality factors an interesting pattern emerged: "High defense pilots" had a poorer performance than "low defense pilots" while at the same time reacted less psychophysiologically (HR and HRV). High defense pilots considered simulator training less important and they trusted more the capability of the aircraft's instrumentation than low defense pilots. As will later be commented, this was contrasted by the fact that high defense pilots (DMT) had more real accidents/'episodes" than low defense pilots.

Lack of motivation did not seem to be a significant factor for performance and activation in this study, since negative correlations were obtained between the degree of realism and difficulty on one hand, and endocrine activation on the other.

Pilots with high DMT DEF were more satisfied with the instrumentation, and at the same time reported less perceived stress and problems. This was interesting since these pilots had more real accidents/ 'episodes' than the others. High scores on general adaptive abilities and psychomotoric performance during pilot-uptake were related to both perceived difficulties and a stronger physiological reaction during the flight.

This should indicate that, independent of psychomotor capabilities, pilots with high defense tended to trust the aircraft's instruments more. They are, therefore, less motivated/interested in different emergency training in the simulator. This can be dangerous since it can reduce the pilot's alertness and increase his reaction time. Such a 'projection' to external factors (to an exaggerated trust in the aircraft) is characteristic for the dynamics of defensivity. Despite a selection based on the NPI (Neumans Pilot Index) on DMT, there are still significant relationships between basic DMT variables such as DMT DEF and Reaction Formation and pilot performance in handling complicated tasks under high stress.

This study confirms previous research which have shown a strong consistent link between defense mechanisms and performance in tasks which are though to demand swift and accurate cognitive performance under conditions of stress, on the one hand. On the other, there are close relationships between the same defenses and both perceived health problems and "physiological stress markers" such as immunoglobuline levels. This has also been found for fighter pilots (16).

What is the potential meaning and significance of these relationships? First, we have empirical evidence that people with high defense strategies tend to have inadequate performance and high autonomic activation in threatening situations. Secondly, such subjects tend not to cope during training, and in the long term develop "burn out" problems. For review see Værnes et al, 1987 (20).

In addition to "defense" as a major concept in our selection research, the theoretical position is based on two major assumptions; activation theory and expectancy. The general activation response is produced whenever ecpectancies are not met. Sustained activation produced psychosomatic disease, and also biochemical changes in the brain which, again relate to poor ability to solve problems.

Expectancy is the property of the brain to store information about the relationship between stimuli and between responses and stimuli. Based on formalisations of these assumptions, a simple system of formal definitions of coping, defense, helplessnesss, and hopelessness has been offered (21). Expectancies may also generalise, this accounts for individual differences in susceptibility to psychosomatic disease.

Multivariate analysis reveal three orthogonal (independent) endocrine factors with specific relations to psychological traits. A catecholamine factor relates to ambition and time urgency, and seems close to the Type A behaviour described as being a cardiovascular risk. A cortisol factor relates to high defense mechanisms. The relations between an androgen and estrogen factor and personality is less stable. When an individual is faced with unsolved problems activation may become sustained and produce pathology through these personality-dependent endocrine reaction systems.

The resulting pathology is not organ specific, but specific with regard to the effector system. For the catecholamine factor, pathology related to the cardivascular system may be expected, but other somatic target organs are also possible. For the cortisol axis, pathology may be produced in several systems. Immunological changes are particularly interesting but may derive also through other systems than the cortisol axis.

It has been shown that tests of psychological defense has proved successful in selecting subjects for a variety of stressful occupations. The level of prediction is many times greater than for other psychological tests which ignore the role of unconscious mental processes. The potential benefits of the DMT are so great that it should be considered seriously for application where tolerance of sudden stress is a key feature.

## REFERENCES


1. *Billings CE*, Reynard WD. Human factors in aircraft incidents: Results of a 7 year study. Aviat Space Environ Med, 55, 1988, 960-965.

2. *Hawkins FH*. Human factors in Flight. Brookfield, Vermont, Technical Press, 1983.

3 *Greenwood M*, Woods HM. A report of the incidence of industrial accidents upon individuals with special reference to multiple accidents. In Hadden W, Suchman EA, Klein D, eds. Accident Research. New York, Harper & Row, 1964.

4. *French JR*, Rodgers W, Cobb S. Adjustment as person-environment fit. In Coelho GV, Hamburg DA, Adams JE, eds. Coping and Adaptation. New York, Basic Books, 1974, 316-333.

5. *Lazarus RC*, Averill J, Opton E. The psychology of coping: Issues of research and assessment. In Coelho G, Hamburg D, Adams J, eds. Coping and Adaptation. New York. Basic Books, 1974, 249-315.

6. *Baade E*, Halse K, Stenhammer PE, Ellertsen B, Johnsen TB, Vollmer F, Ursin H. Psychological tests. In Ursin H, Baade E, Levine S, eds. Psychobiology of stress. A study of coping men. New York, Academic Press 1978, 125-160.

7. *Værnes RJ*. Defensive reactions in dangerous occupations. In Ursin H, Murison R, eds. Biological and psychological basis of psychosomatic disease. New York, Pergamon Press, 1982, 149-162.

8. *Plutchik R*, Kellerman H, Conte HP. A structural theory of ego defense and emotion. In Izard C, ed. Emotion in personality and psychopathology. New York, Plenum Press, 1979.

9. *Kragh U*. The Actual-Genetic Model of Perception-Personality. Lund, Gleerup, 1955.

10. *Murray HA*. The effects of fear on the estimates of the maliciousness of other personalities. J Social Psychol, 4, 1933, 310-329.

11. *Kragh U*. The Defense Mechanism Test: A new method of diagnosis and personnel selection. J Appl Psychol, 44, 1960, 303-309.

*Værnes RJ.* Perceptual defense reactions predict inadequate performance under stress. Scand J Psychol, 23.

13. *Spielberger CD,* Gorsuch RL, Lushene AE. STAI manual for the State-Trait Anxiety Inventory. Palo Alto, Consulting Psychologists Press, Inc., 1970.

14. *Hull CH,* Nie NH. SPSS UPDATE 7-9. New York, 1981.

15. *Nie NH,* Hull CH, Jenkins JG, Steinbrenner K, Bent DH. Statistical package for the social sciences. New York, McGraw-Hill Book Company, 1975.

16. *Værnes RJ,* Myre G, Aas H, Homnes T, Hansen I, Tønder O. Relationships between stress, psychological factors, health and immune levels among military aviators. Submitted, 1988.

17. *Bergan T,* Værnes RJ, Ingebrigtsen P, Tønder O, Aakvaag A, Ursin H. Relationships between work environmental problems and health among Norwegian divers in the North Sea. In Marroni A, Oriani G, eds. Diving and Hyperbaric Medicine. Palermo, EUBS, 1987, 327-336.

18. *Værnes RJ,* Knardahl S, Rømsing J, Aakvaag A, Tønder O, Walther B, Ursin H. Relations between environmental problems, psychology and health among shift workers in the Norwegian Process industry. Work & Stress, 2, 1, 1988, 7-15.

19. *Endresen IM,* Værnes R, Ursin H, Tønder O. Psychological stress-factors and concentration of immunoglobulines and complement components in Norwegian nurses. Work & Stress, 1, 4, 1988, 365-375.

20. *Værnes RJ,* Warncke M, Bergan T, Ursin H. Selecting the right crew for future space stations: An analysis of selection research on offshore divers, aviation pilots and other high risk groups in Scandinavia. Marseille, Proceedings of a Collegium on 'Space & Sea', European Space Agency (ESA), 1987, 47-51.

21. *Ursin H,* Mykletun R, Tønder O, Værnes RJ, Relling G, Isaksen E, Mursion R. Psychological stress-factors and concentrations of immunoglobuline and complement components in humans. Scand J Psychol, 25, 1984, 340-347.

# PERFORMANCE RECOVERY FOLLOWING STARTLE: A LABORATORY APPROACH TO THE STUDY OF BEHAVIORAL RESPONSE TO SUDDEN AIRCRAFT EMERGENCIES

Richard I. Thackray, Ph.D.
Civil Aeromedical Institute
Federal Aviation Administration
P.O. Box 25082
Oklahoma City, Oklahoma, 73125
USA

## SUMMARY

This paper deals with the use of response/recovery to auditory startle as a laboratory technique for simulating some of the principal aspects of the initial shock phase of sudden emergency situations. It is submitted that auditory startle, with its unexpectedness, pronounced autonomic reaction, fear-like subjective experience, and frequent behavioral disruption, approximates the response pattern to be expected in the initial shock phase of sudden traumatic emergencies, and that by studying the time course of performance recovery following startle, as well as individual differences in response/recovery, we may gain a better understanding of some of the variables related to extreme reactions displayed by individuals in real-life emergency situations. Research studies conducted in our laboratory and in others on performance impairment/recovery following startle are reviewed. These studies include those dealing with initial reaction time to the startle stimulus itself, disruption and recovery rate of perceptual-motor (tracking) performance following startle, and the time-course of performance recovery in information processing tasks after exposure to startle. Data are also presented showing a relationship of several individual difference variables to performance response/recovery following startle. These variables include autonomic response to the startle stimulus and level of task proficiency prior to startle.

## INTRODUCTION

Aircraft emergencies often occur without prior warning and require rapid response. Although it is commonly accepted that response times to unexpected events generally exceed those to comparable events that are anticipated, actual data on response times to unexpected stimuli or events occurring infrequently in real-life settings are surprisingly sparse. In one of the few studies in which such data were obtained, Warrick, Kibler, and Topmiller (1965) examined the time that it took secretaries to press a button located 9.5 in from their typewriters when the stimulus (buzzer) was sounded without warning once or twice a week over a 6-month period. Relative to alerted conditions, the increase in response times when the buzzer was unannounced was surprisingly small. During the first month, unalerted response times (Mdn=.8 sec) were about 33 percent longer than response times under alerted conditions. By the end of the 6-month period, the median unalerted time was .6 sec, representing only a 22-percent increase over alerted times.

Other studies of response times to unexpected events have been conducted by investigators concerned with driver reactions to simulated emergencies. Muto and Wierwille (1982), for example, found that braking time to an unexpected event, presented after prolonged driving, averaged about 1.64 sec when the event first occurred. By the time the fourth "emergency" occurred, response times were about equal to baseline response times (approximately 1.40 sec). Thus, unexpectedness resulted in braking times that were 23 percent longer, at most, than braking times when the events were anticipated. In a somewhat similar study, Johansson and Rumar (1971) also compared braking response times to expected and unexpected situations. On the average, braking time to unexpected situations averaged .73 sec; this decreased to .54 sec when the events were anticipated. Unexpectedness, thus, resulted in response times that were approximately 35% longer than response times for anticipated events.

A few reported studies have dealt with simulated nuclear power plant emergencies. In these studies, process operators in nuclear control rooms were instructed to respond as rapidly as possible to simulated emergencies signaled by audible alarms and visual indicators. With signal rates of 1.35 to .35 per hour, response times (estimated from the data given) ranged from less than 1 sec to approximately 2.5 sec (Lees and Sayers, 1976).

Of the studies just discussed, those that have compared response times to both expected and unexpected stimuli are relatively consistent in their findings. Maximum percent increase in response time due to the factor of unexpectedness has been found to range from 22 to 35 percent. When the influence of repetition has been examined, reduction in uncertainty caused response times to approximate baseline (alerted) conditions. Such findings lend support to the conclusion reached by Warrick, Kibler, and Topmiller that one may be able to extrapolate to unalerted conditions from data collected under comparable alerted conditions.

In many types of emergency situations, however, one has not only the factor of unexpectedness to contend with, but also the additional and potentially disruptive factor of intense emotional arousal. Actual data with regard to response time to traumatic emergency events, to say nothing of the time-course of behavioral recovery following such experiences, are virtually nonexistent. Part of this is clearly due to the extreme difficulty of creating under controlled, experimental conditions the particular perceptual/cognitive events that, because of their meaning or significance to the individual, are the usual triggers for the emotional reactions associated with real-life emergencies.

RATIONALE FOR THE USE OF STARTLE

A possible technique for circumventing this dilemma involves the use of startle. Before considering this approach, however, a brief review of the startle response is warranted. In essence, the startle reflex is primarily a muscular response where the complete reaction consists of a series of involuntary contractions beginning at the head with the eyeblink and rapidly progressing to the legs. It is typically evoked by impulsive auditory stimuli (e.g., a pistol shot), although other, and generally less effective stimuli, such as a jet of ice water, photoflash, and electric shock have also been found to elicit it (Landis and Hunt, 1939). It always begins within 100 msec of the eliciting stimulus, and may have a duration of .3 sec for a mild but complete response to approximately 1 to 1.5 sec for an intense reaction (Ekman, Friesen, and Simons, 1985; Landis and Hunt, 1939). Although the muscle reflex, described in detail by Landis and Hunt (1939), is often considered to define the startle pattern in its entirety, the total pattern includes physiological as well as subjective components. The physiological response consists of a pronounced, generalized increase in autonomic and central nervous system activity and has been described in detail by Sternbach (1960a). This pattern of physiological response, when compared with autonomic response patterns produced by exercise, the cold pressor test, and injections of epinephrine and norepinephrine, has been found to closely resemble the pattern produced by epinephrine injection (Sternbach, 1960b).

The feeling state evoked by startle is more difficult to classify. While often considered to be related to the emotion of surprise (Ekman, Friesen, and Simons, 1985), others have identified it not only with surprise, but with fear and anger as well (Blatz, 1925; Landis and Hunt, 1939; Skaggs, 1926). Interestingly enough, the epinephrine-like physiological pattern to startle that was noted above is also the characteristic pattern found to be produced by fear-inducing situations (Ax, 1953; Schacter, 1957). Although agreeing that the feeling state associated with startle appears closest to fear and anger, Landis and Hunt (1939) consider that it may be best to define startle as preemotional. They note that "It does not stand in the same group of phenomena as the major emotions, yet it seems to be closely related to them and to belong generically in the same field. It is an immediate reflex response to sudden, intense stimulation which demands some out-of-the-ordinary treatment by the organism. As such it partakes of the nature of an emergency reaction, but it is a rapid, transitory response much more simple in its organization and expression than the so-called 'emotions'" (Landis and Hunt, 1939, p. 153).

In a study concerned with the question of why some individuals seem to "freeze," while others appear to react almost instantaneously in emergency situations, Sternbach (1960a) reasoned that startle resulting from a loud auditory stimulus might be used to approximate the principal components (surprise, fear, intense physiological arousal, and temporary behavioral disruption) that are common to many types of sudden emergencies and hence provide a technique for studying behavioral recovery following traumatic events under laboratory conditions. It is generally accepted that sudden emergencies frequently, if not typically, elicit feelings of fear or anxiety, and, as we have just noted, a number of studies have demonstrated that startle does evoke an experience, albeit rather transitory, that has been identified not only with surprise, but with fear as well. Further, the physiological response to startle, when compared with the autonomic response patterns produced by a number of other stressors, has been found to closely resemble the epinephrine pattern associated with fear-inducing situations. Taken in conjunction with the Landis and Hunt (1939) belief that the total startle pattern resembles that of an emergency reaction, it would not seem unreasonable to believe that studies of response to startle might provide a useful laboratory approach to the study of human behavior in sudden stress situations. The present paper adopts this position and reviews research findings relevant to performance recovery from startle. No attempt is made here to document the methodological considerations (e.g., stimulus parameters, modifying variables, differentiation of startle from orienting and defensive reflexes, measurement requirements) that must be recognized in carrying out research in this area. Relevant methodological considerations are reviewed or described by Graham, 1979; Landis and Hunt, 1939; Ekman, Friesen, and Simons, 1985; Raskin, Kotses, and Bever, 1969, and Thackray, 1972.

RESPONSE TIME TO STARTLE

Using a pistol shot as the stimulus for a required button press response, Sternbach (1965a) found that voluntary response times to startle stimulation ranged from 128 to 3,262 msec with a mean (estimated from the data) of 950 msec. Sternbach's primary concern, however, was not with establishing the actual range or limits of response time to startling events, but rather with investigating psychophysiological correlates of individual differences in time to respond. In this regard, he examined physiological resting and response levels of the 10 fastest and slowest reactors to startle. While there was no meaningful relationship of resting physiological levels to reaction time, fast and slow reactors differed significantly in their physiological response to startle on a number of variables; slow reactors showed a significantly greater increase in systolic blood pressure, pulse pressure, palmar skin conductance, and heart rate than did fast reactors. In addition to greater autonomic response, informal statements made by slow reactors (e.g., "I knew I was supposed to do something, but I couldn't think of it at first." "I thought I pressed it at first, then I realized I hadn't." "It took me a moment to realize what I had to do.") suggested greater cognitive disruption as well; no such statements were made by the group of fast reactors.

A subsequent study by Thackray (1965) extended the Sternbach study by including a comparison of response times to high-intensity, startling stimuli with reaction times to nonstartling auditory stimuli. The principal intent of this investigation was to provide baseline data that might be used to estimate pilot response times to potentially critical situations, such as unexpected clear air turbulence or a sudden failure in an automatic control system. Subjects were instructed to respond to any auditory stimulus by moving a control stick as rapidly as possible to the left and simultaneously flipping back a response button located on top of the stick. The first stimulus consisted of an unexpectedly loud burst of 120-db noise; this was followed by a series of 50 low-intensity auditory stimuli at constant 15 sec intervals and a final 120-db stimulus. The mean (893 msec) and range (356 to 1800 msec) of response times to the initial high-intensity stimulus were similar to those obtained by Sternbach. Like Sternbach, autonomic reactivity to startle was found to be positively correlated with response time to startle. The second high-intensity stimulus presented 15 sec after the series of low-intensity stimuli, and with no indication that anything other than another low-intensity stimulus would occur, yielded a mean (416 msec) and range (187 to 1550 msec) of response times that were considerably lower than that obtained to the first high-intensity stimulus. Interestingly enough, autonomic response to the second loud stimulus was found to be inversely related to response time. Thus, while magnitude of autonomic response to the initial high-intensity sound was directly related to performance disruption, autonomic response to the second, and subjectively less startling sound, was associated with performance facilitation. One might hypothesize that, in accordance with the predictions of activation theory (Malmo, 1959), arousal level to the initial startle was sufficiently high to disrupt performance, while the lower arousal associated with the second startle acted to facilitate performance.

Although positive correlations were found between reaction times to the low-intensity sounds (Mn=368 msec) and response times to the high-intensity, startling stimuli, the most interesting aspect of this finding was that startle appeared to magnify differences between individuals in their reaction times to the low-intensity, nonstartling tones; i.e., slow responders tended to respond even more slowly, while the fast responded more rapidly to startle stimulation.

RESPONSE/RECOVERY OF CONTINUOUS PSYCHOMOTOR PERFORMANCE FOLLOWING STARTLE

While the studies described above provide basic information on the time required to make a discrete, voluntary response to startle, they fail to indicate whether this time frame encompasses all of the disruptive effects of startle or whether some disruption may extend beyond this period. Since the reflex muscle response to startle, depending upon the intensity of the reaction, may last from .3 to 1.5 sec (Landis and Hunt, 1939), it is evident that a major portion of the time required to complete a voluntary response following startle is a direct result of this reflex interference. To provide information on possible disruptive effects of startle beyond this period, Thackray and Touchstone (1970) studied the recovery rate of continuous psychomotor performance following startle. In this study, subjects performed a compensatory tracking task continuously during a 30-min period. A 115-db burst of white noise occurred unexpectedly 2 min into the session and again at the middle of the session. Tracking error during the first minute following the initial startle stimulus is shown in Figure 1. Also shown in this figure are the response/recovery curves for heart rate and skin conductance. Although maximum performance disruption occurred during the first 5-sec measurement period following stimulation, significant ($p<.05$) impairment was still present 10 sec after startle.

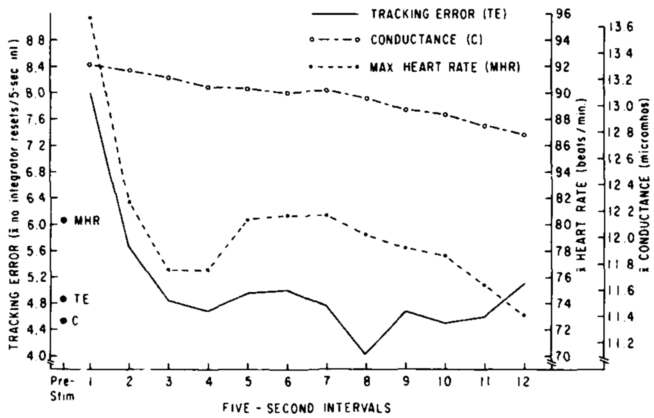



Figure 1. Mean tracking error, maximum heart rate, and conductance level during successive 5-sec intervals following startle. Also shown are pre-startle values for each variable.

The disruption in tracking performance, persisting into the 10-sec period following startle stimulation, clearly extended beyond the initial disruption caused by the reflex response itself and would appear to be a manifestation of a longer lasting, more general physiological/emotional response to the unexpected noise stimulation. Support for this view is suggested by the apparent covariation of heart rate with performance that is shown in Figure 1 and that appears to extend at least into the first 30 sec following stimulation. (Incidently, it is of interest to note in this figure that significant performance <u>improvement</u> occurred during the 8th 5-sec interval following startle; facilitation at this same location also occurred following the second startle stimulus. Since neither of the autonomic measures showed any corresponding change during this time period, some central nervous system facilitory process is suggested.)

The pattern of performance change and physiological response to the second of the two startle stimuli, although of somewhat lower magnitude, was quite similar to that shown in Figure 1. Of interest was the finding that magnitude of tracking error to the two startle stimuli was significantly correlated ($r=.60$, $p<.01$). This enabled us to form two subgroups of subjects whose tracking error following both startle events placed them in either the top third (high impairment) or bottom third (low impairment) of the combined distributions. Relative to prestartle tracking performance, it was found that the high-impairment group almost doubled in their tracking error scores immediately following startle; the low-impairment group showed little difference between their prestartle and poststartle levels of tracking error. With regard to physiological response to startle, the high-impairment group showed significantly greater heart rate acceleration, but the groups did not differ significantly ($p>.05$) in conductance change.

A study by Vlasak (1969) likewise evaluated individual differences in psychomotor disruption to startle stimulation. Using a simple line-tracing task, Vlasak studied differences in performance disruption to an unexpected 100-db sound from a Klaxon horn. His findings were similar to those of Thackray and Touchstone (1970); performance impairment following startle was related to prior task proficiency, with less proficient subjects being considerably more disrupted by startle. As noted earlier, Thackray (1965) also found evidence to suggest that, with the particular reaction time task employed, startle tended to exaggerate preexisting differences between individuals in their nonstartle response time; i.e. the slow became slower and the fast responded with even shorter latencies to startle. Taken together, the results of these three studies suggest the general hypothesis that the extent of disruption following startle is dependent upon prestartle level of performance, with the greatest impairment occurring among those who are either slowest or least proficient prior to startle.

Before concluding this section it should be noted that both Vlasak and a subsequent study by May and Rice (1971) found the total duration of tracking impairment following startle to be only 2 to 3 sec, which is considerably less than that found in the Thackray and Touchstone study. In a reexamination of their data, Thackray and Touchstone likewise found maximum impairment to occur within this same time period and concluded that at least some of the disruption that takes place within the 5-sec period following startle is attributable to direct mechanical effects of the muscle reflex on motor control. However, the fact that Thackray and Touchstone found tracking performance to be significantly impaired for up to 10 sec following startle clearly demonstrates that disruptive effects transcend the time period that one might reasonably attribute to mechanical effects of the startle reflex. The longer period of disruption found by Thackray and Touchstone may have been due to the use of a more difficult tracking task and/or the use of a more refined measure of tracking error than was used in either the Vlasak or the May and Rice study.

RECOVERY OF COGNITIVE FUNCTIONING FOLLOWING STARTLE

Although perceptual-motor recovery following startle appears to be quite rapid, there is evidence that tasks involving decision making or information processing may be impaired for a longer period of time. Thus, Vlasak (1969) studied the effects of startle on continuous mental subtraction and found performance to be signficantly impaired during the first 30 sec following stimulation. A similar period of impairment was found by Woodhead (1959, 1969), who obtained decrements on a continuous symbol-matching task lasting from 17 to 31 sec after startle. The fact that impairment on some tasks following startle may last for at least 30 sec lends further support to our belief that startle effects may extend considerably beyond the initial period of motor disruption produced by the reflex response itself.

In all of the startle studies just reviewed, however, performance recovery effects were studied only during some portion of the first 60 sec following stimulation. While it is certainly possible that performance impairment does not extend beyond this time period, startle is known to be accompanied by rather pronounced autonomic (especially cardiovascular) changes (e.g., Thackray and Touchstone, 1970, 1983), and it is conceivable that such changes could have more lasting effects on performance. Thus, a pronounced discharge of the autonomic nervous system might have a long-term activating effect leading to performance facilitation, or, conversely, it might produce a period of parasympathetic overcompensation resulting in eventual drowsiness and impaired performance.

In our most recent study (Thackray and Touchstone, 1983), we used monitoring and information processing tasks to examine both short- and long-term performance recovery effects following a simulated emergency situation (a radar failure) that was accompanied by either a startling or a nonstartling auditory signal. The subject's primary task was to monitor a simulated air traffic control (ATC) radar display. One hour into the session a radar failure occurred that was accompanied by either a loud (104 db) or low level (67 db) burst of white noise acting as an alarm signal. Subjects were then required to turn in the chair and begin performing a simple information processing (serial reaction) task. (The serial reaction task consisted of a self-paced, four-choice reaction time task in which the subject pressed one of four keys in response to a centrally displayed number.) Five minutes of performance on this task was followed by a return to radar monitoring. In addition to performance, physiological and subjective measures of startle and arousal were also obtained. It was hypothesized that performance following the high-intensity alarm signal (expected to elicit a startle reflex) would be significantly impaired relative to performance following the low intensity signal (expected to elicit an orienting-type response).

Heart rate response and subjective ratings of startle were consistent in demonstrating that the high-intensity signal was clearly startling to subjects in this group. Conversely, the group exposed to the low-intensity signal did not rate the signal as startling, and the slight heart rate deceleration that occurred immediately following stimulation was consistent with the expectation that this level of noise would produce only an orienting or surprise reaction (Graham, 1979). In spite of these differences, however, both groups showed almost identical patterns of performance change during the first minute following noise stimulation. Relative to prestimulus levels, mean response times on the serial reaction (SR) task were significantly elevated only during the first 6 sec following noise; thereafter, performance returned to prestimulus levels for the remainder of the 60-sec period. A comparison of the response patterns obtained for the two groups is shown in Figure 2.

At first glance, this lack of any difference between the startled and nonstartled groups in mean performance during the first 6 sec following stimulation would appear to be inconsistent with the findings of our previous studies and those of others reviewed earlier. Since these results were not expected, response times during the first 6-sec period were examined more closely. The time from the onset of the noise signal to the first SR response was obtained for each subject. These initial SR response times, which encompass the time required to transition from the radar to the SR task, were plotted on log probability paper and are shown in Figure 3. Although mean time to make this initial response (designated task transition time) did not differ among the two groups (2.91 and 2.84 for the means of the high- and low-intensity groups respectively), Figure 3 clearly suggests a difference between the groups in range or variability of transition times. An F test of the variances of the two groups revealed the startled group to be significantly more variable ($F(14/14)=2.61$, $p<.05$) in the time required

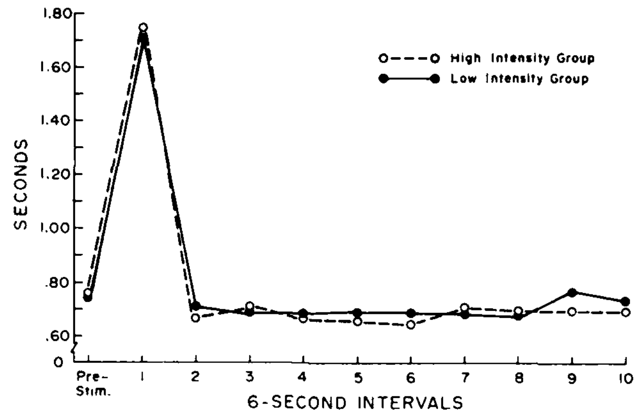



Figure 2. Mean response time for SR performance during successive 6-sec intervals of the first minute following noise stimulation. Also shown are prestimulus values.

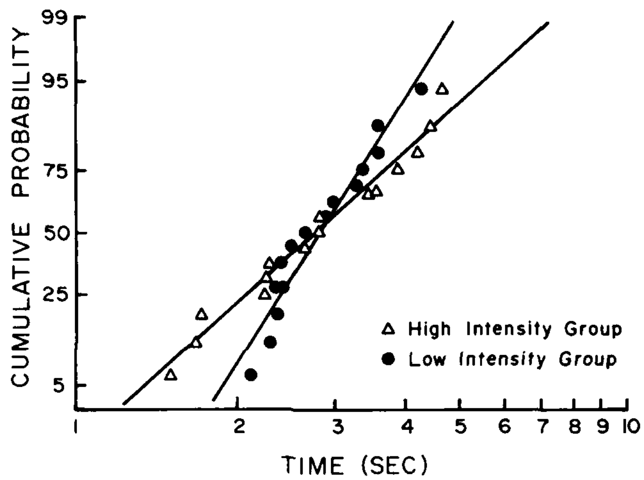



Figure 3. Task transition times for the two groups.

to make this initial response. An examination of variability of responses on the SR task subsequent to this first response, but still within the first 6-sec period following stimulation, revealed variances of .2869 and .1272 for the high- and low-intensity groups respectively. These values, although in the same direction as the transition time variances, failed to reach significance ($F(14/14)=2.25$, $p>.05$). The difference between groups in response variability was thus confined to task transition time.

Analyses of the video-taped recordings taken during noise stimulation clarified these findings. In the group receiving the nonstartling noise signal, behavior following stimulation was extremely uniform; subjects slowly turned in the chair and began performing the SR task. In the high-intensity (startle) group, there were pronounced individual differences following stimulation with some subjects appearing dazed and confused by the noise while others recovered almost immediately and rapidly began performing the task. The disruptive effect of the loud sound for some subjects combined with the rapid recovery shown by others apparently balanced the generally uniform response of the low-intensity group. This also explained the difference in the variance of response times of the two groups. The increased range or variability of initial response to startle that was found in this study is clearly similar to that discussed earlier in the context of both voluntary reaction time to startle and tracking performance following startle.

Unlike response times which, except for the initial task transition time, were largely unaffected by startle, the frequency of incorrect responses (representing errors in information processing) was found to be significantly greater in the startled than in the nonstartled group during the first minute following stimulation. This finding is in general agreement with the findings of Vlasak (1969) and Woodhead (1959, 1969) mentioned earlier, that information processing may be impaired during recovery from startle for periods ranging from 17 sec to over 30 sec. Woodhead (1969) has noted that 30 to 60 sec is the period that it generally takes for autonomic responses such as heart rate to recover to approximate prestimulus levels following startle and that it may not be mere coincidence that this corresponds to the recovery period of cognitive performance.

There was no evidence that startle affected frequency of errors or mean performance on either the SR task or on the radar task subsequent to the first minute following stimulation. Since neither heart rate nor conductance level differed among the groups during these subsequent periods of SR and radar performance, it may be concluded that both the physiological and performance effects of startle are largely confined to the initial 1-min period following startle stimulation.

FIELD STUDIES OF RESPONSE/RECOVERY TO STARTLE

It would be desirable to compare laboratory findings of performance recovery from startle with the findings of comparable studies conducted in the field. Unfortunately, such comparisons are few because of the paucity of published findings. In one of the few field studies of which I am aware that specifically investigated the effects of startle on performance, Ziperman and Smith (1975) compared the extent of disruption of driving behavior produced by unexpected air-bag deployment with that resulting from hood fly-up. Fifty-one male and female drivers ranging in age from 19 to 74 years were tested. Although air-bag deployment, accompanied by a shot-like sound, was experienced as being considerably more startling than hood fly-up, both types of events produced similar, marked changes in heart rate, blood pressure, and skin conductance. In spite of pronounced subjective and physiological evidence of startle, drivers apparently retained control of the test vehicle and were reported to be lucid on questioning less than 10 seconds after cushion deployment. As stated in their paper, "The average steering-wheel rotation was 85 degrees during hood fly-up and 72 degrees during cushion deployment. This degree of steering-wheel rotation would correspond to approximately 3 to 4 degrees at the tire. In combination with the lateral-deviation data, it shows that adequate steering control can be and is maintained in the startle modes tested" (p. 439). Although the effects of these startling events might appear to be less than one might have expected, it should be noted that the actual time-course of performance recovery cannot be determined from the data as reported in this study. There is no indication, however, that the duration of performance disruption found by Ziperman and Smith would differ appreciably from that found in our laboratory studies.

CONCLUSIONS

If we combine the results of all studies considered thus far, certain generalizations concerning response/recovery following startling events can be made:

1. Simple, voluntary responses to startling stimuli or events can generally be made within 1 to 3 sec following stimulation (Sternbach, 1960a; Thackray, 1965; Thackray and Touchstone, 1983). In this regard, mean time to respond to a startling stimulus may not differ appreciably from mean time to respond to an unexpected event or stimulus that is simply surprising. It is likely, however, that the range of response times to the former type of event will significantly exceed the range of response times to the latter type of event (Thackray, 1965; Thackray and Touchstone, 1983).

2. More complex perceptual-motor behavior, such as that requiring continuous psychomotor control, is likely to show maximum disruption during this same 1- to 3-sec period (May and Rice, 1971; Thackray and Touchstone, 1970, 1983; Vlasak, 1969; Ziperman and Smith, 1975), although significant, but lesser, disruption may still be present for up to 10 sec following stimulation (Thackray and Touchstone, 1970).

3. Evidence from several studies suggests that the ability to process information may be impaired for 17 to 60 sec following the occurrence of a startling event (Thackray and Touchstone, 1983; Vlasak, 1969; Woodhead, 1959, 1969).

4. Individual differences in the magnitude of performance impairment following startle appear (a) directly related to physiological reactivity to startle (Sternbach, 1960a; Thackray, 1965; Thackray and Touchstone, 1970) and (b) inversely related to level of prestartle task proficiency (Thackray, 1965; Thackray and Touchstone, 1970; Vlasak, 1969).

In order to evaluate the relevance of the above laboratory and field findings of response/recovery following startle to behavioral response following real-life emergencies, it is important to recognize that unexpected and traumatic emergency situations in real life probably involve at least two phases. The first phase, which could be termed a "shock phase," constitutes the initial reaction. In this phase, the individual attempts to respond with immediate behaviors that are intended to cope with or rectify the unexpected event. While the behaviors employed may appear to be irrational and actually worsen the situation, this is clearly not the intent. With some individuals, behavior seems to become suspended (affective immobility or "freezing"), although numerous studies of response to disaster (e.g., Singer, 1982) suggest that this type of response is the exception rather than the rule. When it does occur, it appears to be a rather temporary or momentary response. In some emergencies, the shock phase is followed by a second phase which could be termed an "evaluative phase." This phase occurs if the emergency situation has not been resolved during the intial shock phase and is characterized by an emerging perception or evaluation of the situation in terms of the individual's ability, or lack of ability, to cope with the emergency. It is during this phase that panic, if no solution or escape seems possible, may occur. However, panic, like affective immobility, also appears to be a relatively infrequert form of disaster response (Singer, 1982).

If one is willing to accept that the emotional/physiological response to startle can serve to at least approximate the initial shock phase of traumatic, real-life emergencies, then findings of laboratory studies of performance recovery following startle may have relevance in predicting the time course of behavioral recovery following such events and may assist in our understanding of some of the extreme reactions displayed by individuals in real-life emergency situations. As we have noted, laboratory studies have isolated several individual difference variables (autonomic reactivity and level of prior task proficiency) that appear to be correlated with performance recovery from startle. The first of these, autonomic reactivity, suggests that inherent, constitutional factors undoubtedly play some role in startle recovery; the second variable, task proficiency or skill level, would suggest that some of the performance disruption following startle may be amenable to training. Research is needed, however, to determine the extent to which individual differences in response/recovery found in laboratory studies of startle can serve as useful predictors of disruption/recovery following simulated emergencies that closely approximate real-life situations.

## REFERENCES


1. Ax, A. A. The physiological differentiation between fear and anger in humans. Psychosom. Med., 15, 1953, 433-442.

2. Blatz, W. E. The cardiac, respiratory, and electrical phenomena involved in the emotion of fear. J. Exp. Psychol., 8, 1925, 109-132.

3. Ekman, P., W. V. Friesen, and R. C. Simons. Is the startle reaction an emotion? J. Pers. Soc. Psychol., 49, 1985, 1416-1426.

4. Graham, F. K. Distinguishing among orienting, defense, and startle reflexes. In H. D. Kimmel, E. H. van Olst, and J. F. Orlebecke (Eds.), The Orienting Reflex in Humans. New York: Erlbaum Associates, 1979, 137-167.

5. Johansson, G. and K. Rumar. Drivers' brake reaction times. Hum. Factors., 13, 1971, 23-27.

6. Landis, C. and W. A. Hunt. The startle pattern. New York: Farrar and Rinehart, 1939.

7. Lees, F. P. and B. Sayers. The behavior of process operators under emergency conditions. In T. B. Sheridan and G. Johannsen (Eds.), Monitoring Behavior and Supervisory Control. New York: Plenum Press, 1976.

8. May, D. N. and C. G. Rice. Effects of startle due to pistol shots on control precision performance. J. Sound Vib., 15, 1971, 197-202.

9. Malmo, R. B. Activation: A neuropsychological dimension. Psych. Rev., 66, 1959, 367-386.

10. Muto, W. H. and W. W. Wierwille. The effect of repeated emergency response trials on performance during extended-duration simulated driving. Hum. Factors., 24, 1982, 693-698.

11. Raskin, D. C., H. Kotses, and J. Bever. Autonomic indicators of orienting and defensive reflexes. J. Exp. Psychol., 80, 1969, 423-433.

12. Schacter, J. Pain, fear, and anger in hypertensives and normotensives. Psychosom. Med., 19, 1957, 17-29.

13. Singer, T. J. An introduction to disaster: Some considerations of a psychological nature. Aviat. Space Environ. Med., 53, 1982, 245-250.

14. Skaggs, E. B. Changes in pulse, breathing and steadiness under conditions of startledness and excited expectancy. J. Comp. Psychol., 6, 1926, 303-318.

15. Sternbach, R. A. Correlates of differences in time to recover from startle. Psychosom. Med., 22, 1960a, 144-148.

16. Sternbach, R. A. A comparative analysis of autonomic responses to startle. Psychosom. Med., 22, 1960b, 204-210.

17. Thackray, R. I. Correlates of reaction time to startle. Hum. Factors., 7, 1965, 74-80.

18. Thackray, R. I. Sonic boom exposure effects II.3: Startle responses. In R. Rylander (Ed.), Sonic boom exposure effects: Report from a workshop on methods and criteria, Stockholm 1971. J. Sound Vib., 20, 1972, 477-544.

19. Thackray, R. I. and R. M. Touchstone. Recovery of motor performance following startle. Percept. Mot. Skills., 30, 1970, 279-292.

20. Thackray, R. I. and R. M. Touchstone. Rate of recovery and subsequent radar monitoring performance following a simulated emergency involving startle. FAA Office of Aviation Medicine Report No. FAA-AM-83-13, 1983.

21. Vlasak, M. Effect of startle stimuli on performance. Aerosp. Med., 40, 1969, 124-128.

22. Warrick, M. J., A. W. Kibler, and D. A. Topmiller. Response time to unexpected stimuli. Hum. Factors., 7, 1965, 81-86.

23. Woodhead, M. M. Effect of brief noise on decision making. J. Acoust. Soc. Am., 31, 1959, 1329-1331.

24. Woodhead, M. M. Performing a visual task in the vicinity of reproduced sonic bangs. J. Sound Vib., 9, 1969, 121-125.

25. Ziperman, H. H. and G. R. Smith. Startle reaction to air-bag restraints. J. Amer. Med. Assn., 233, 1975, 436-440.

# PERIPHERAL NERVOUS VELOCITY OF CONDUCTION IN FIGHTER PILOTS

Major,J.L.Garcia Alcón,MC,SAF,Flight Surg. at 23th Wing. TALAVERA AFB.
Professor,J.M.Moreno Vázquez,MD,Department of Physiology. Extremadura University.
Senior Prof.,J.E.Campillo Alvarez,MD,Dept. of Physiology. Extremadura University.
Dr.,A.Gonzalez Roncero.Clin.Neurophysiol.National Health Institute.
Postal adress: Avda Ramón y Cajal 5-06001-BADAJOZ-SPAIN

## SUMMARY

Fighter pilot is under an important stress because of his special professional activity. It origins an automatic response through neurohormonal mechanisms.The most important - among these mechanisms is the catecholamins secretion.These hormones will produce very important changes in the general homeostasy.The peripheral nervous system and mainly its mielyn sheath,is highly sensitive to variations in the internal environment.When that sheath is damaged the rapidity of nervous impulse transmission decrease.The system for to know that injury,is through measuring of the nervous velocity of conduction. This work shows the abnormal behaviour of sensitive nervous velocity of conduction in fighter pilots in depending of flight hours.The great consumption of oxygen could be - the reason for that alteration.Authors have found an important increase in Catalase and Glutathione-Peroxidase,that enzymes are protective systems in front to oxidations.

Fighter pilot is under an important stress because of his special professional - activity.This stress is developed in the psichologycal context as well as in the physical field.We will discuse our work only focusing on some aspects of these physical stress.When the organism has been exposed to stress,it origins an automatic response - using a reaction in its neurohormonal mechanisms.The most important among these mechanisms is the catecholamins secretion.These hormones will produce very important changes in the general homeostasy such as an increase in anaerobic activity with the resulting production of acid metabolites (12).

The peripheral nervous system and mainly its mielyn sheath,is highly sensitive to variations in the internal environment.As a typical examples neuropathie of Diabetes Mellitus (7,9,14) and Uremic Syndrom (3) as well as another toxic syndroms have a clear and objective manifestation which is a decrease in the nervous velocity of conduction. According to Jumping Theory which is related to nervous conduction,the potential of action will travel along the axon and will jump from Ranvier's nodule to Ranvier's nodule. Those areas have not mielyn sheath and so, the nervous impulse runs in a faster and easier way. Therefore when mielyn sheath is damaged the rapidity of nervous impulse - transmission decreases.The system for to know that injury, is through measuring of the nervous velocity of conduction.

## AIMS

One: As the nervous system is extremely sensitive to every kind of agression,we wanted to objectivize the flight stress repercussion over it.

Two: Afterwards we have attempted to correlate the information obtained with - flight hours number of each pilot and consequently with his age.

Three: Whereas the results obtained in previous points,and suspecting that their were caused by oxygen excess inhaled for pilots in their professional activity daily and that oxygen stress is one among a few stress factors in flight,we wanted to valuate the enzymatic repercussion due to increase of free radicals resulting from an oxidation increase.

## METHODS

This study has been performed with 21 volunteers of flight and attack pilots instructors course with a flight experience between 300 flight hours - the youngers -and 2600 hours- the olders.All of the hours mentioned were flown in jet aircraft.Because- of methodologycal reasons we shaped two groups: A.- Made with pilots aged from 21-30 years old (n=9),and B.-Made with pilots aged from 31-40 years old (n=12).We performed the constrols according to normality tables obtained in our laboratory (13),based upon a widw study on healthy subjects allocated in diferentes ages and sexes.We have been rather rigorous in the making of the groups because of the proved repercussion of - this factors on peripheral nervous velocityof conduction. See Table I.

The neurophysiologycal study has been performed with a Evoked Potentials advice, (VSC6-MEDELEC) with one stimulator unit and a digital average (DAV6-MEDELEC).

Enzymatic study has been developed according to standard technics by spectrophotometry through methods shown by AEBI (for Catalase) and FLOHE and GUNZLER (for Glutathione-Peroxidase),(1,4).

We obtained sensitive conduction velocities by stimulating the cutaneous area - corresponding to nerve under exploration.For this purpose two ring electrodes,(anode- and cathode),were placed in the fifth finger for ulnar nerve and in the third finger for median nerve. Cathode always was placed in proximal situation.The evoked potential was pick up at the ventral side of the wrist.This was made by placing percutaneous e-- lectrodes at the cubital side or de middle side according to nerve to be explored.Cathode is always in distal situation.The distance between the stimulation area and the collecting area is measured from one cathode to the other (5),see figure 1.We read the latency directly in osciloscopy screen by digital cross slide.Motor nervous velocity of conduction was obtained by similar technique except in the situation of electrodes.

The placement of electrodes for stimulation was at the ventral side of wrist and elbow. To pick the response up we placed percutaneous electrodes at the short abductor muscle of the first finger, when we explore median nerve. (Figure 2).Afterthats latency diferences are calculated.

## RESULTS

In front of normal decrease of the peripheral sensitive nervous velocity of con-- duction according to age, we don't find that decrease in the pilots studied. Nevertheless, we observe in median nerve a paradoxical increase of velocity with aging.The evolution of the motor conduction is the same as in the control group, but their values / are significantly higher. (In all results $p < 0.01$).(Figures 3 and 4)

Later on, when we correlated the velocity of conduction with the flight hours experience, we found a significant increase in sensitive velocity of conduction at median nerve in relation to the flight hours experience. We didn't find a increase in other / nerves. (Figures 5 and 6)

We calculated at median nerve correlation coefficient and regression line in order to ensure positive correlation between sensitive velocity of conduction and flight hours. From both correlation coefficient and regression line, we obtained a formula,relating both velocity of conduction and flight hours, mathematically.(Figure 7).

Supposing those modifications in velocity of conduction to be subordinate on increase of oxygen consumpltion in the pilots and taking LOW's Hypoxic Theory in the diabetic neuropathy like reference, we determined some protective enzymes in front of oxydations. We observe an important increase of Catalase and Glutathione-Peroxidase very overhead normal range in control groups. ($p < 0.01$). (Figure 8)

## DISCUSSION

Majority of authors (2,8) agree to point out that sensitive velocity of conduction as a tendency to decrease depending of the age, while motor velocity of conduction as a tendency to remain the same,with small oscilation. We made a wide study about the influence both sex and age over nervous velocity of conduction,obteining tables that,in general line, are similar to the other authors. This tables are current references for us in our investigation about repercussion of diferent nosological entities over nervous / velocity of conduction.

When we compare the sensitive velocity values in two problem groups (A and B) with the respectives control groups (AC and BC), we observe an abnormal behaviour,above mentioned. Seeing that the accumulation of flight hours is the only commun factor in all pilots,with aging, we correlated this date with velocity of conduction, obtaining the same results that in prior comparation. The positive correlation between flight hours and velocity of conduction were mathematically demostrated by means of correlation coefficient calcule, as we said previously.

LOW and colaborators (9,10,11) to study phisiopatology of Diabetic Neuropathie, suggest that deficit of oxygen into nervous tissue provoke this entity, enunciating the Hypoxic Theory. This propose that the diabetic state, by unknown mechanisms, results in rheologic an capillary abnormalities that lead to decreased blood an oxygen delivery to nerve, resulting metabolic processes and a vicious cycle of scalating microvascular damage and further hypoxia. Chronic hypoxia results in slowing of nerve conduction and resistance to ischemic conduction block. Oxygen suplementation also result in normalization of velocity of conduction. According to,the decreasing peripheral nervous velocity of conduction, is the most usual finding in that complication of the Diabetes Mellitus. From this theory, LOW defends to oxygenotherapy as basic treatment of mentioned neuropathie, obtaining recuperation in velocity of conduction.

Extrapolating our results to LOW's Hypoxic Theory, an searching an explication to abnormal evolution at sensitive velocity of conduction in the pilots, we supposed that alteration could root in that pilot during their professional activity, they are expose to greater comsumption of oxygen than the rest of the people.

Oxygen supplied at concentrations greater than those in normal air as long being know to be toxic to plants, animals, and to aerobic bacterias. The toxicity of oxygen to animals, including Man, as been of interest in relation to diving, underwater simming and scape from submarines, and more recently, in the use of oxygen in the treatment of the cancer, gas gangrene, multiple sclerosis and lung diseases, and in the design of the gas supply in spacecraft. High oxygen concentration also cause a general "stress / reaction" in animals, with stimulates the action of some endocrine glands.

GERSHMAN and GILBERT, in the USA, to propose in 1954, that most of the damaging effects of oxygen, could be atributed to the formation of free oxygen radicals. What is a free radical ? According to HALLIWELL (6) a free radical is "any species capable of independent existence that contains one or more unpaired electrons". The presence of that electrons causes the species to be attracted slightly to a magnetic field an sometimes makes the species highly reactive. Superoxide radical is the last product in / respiratory chaine, as was first observed by FENTON in 1894: a mixture of hidrogen peroxide and iron salt, reacts and originate in two times Superoxide and Hidroxil radicals. Thus the Fenton's reaction is a biological reality if hidrogen peroxide is available in vivo as well as the respiration process.

Human organism has a lot of protective mechanism in front of the increase of oxidations into the cell, by means of enzymatic systems activation. This sysstem remove the free radicals, that have its origen in oxidative reactions from cells. Two types of enzymes exist to remove hidrogen peroxide within cells: They are the Catalases which catalyse reaction ($2H_2O_2 \rightarrow 2H_2O + O_2$) and the Peroxidases, which bring about the general reaction ($SH_2 + H_2O_2 \rightarrow S + 2H_2O$) in which $SH_2$ is a substrate that becomes oxidized.In man, both enzymes Catalase and Glutathione-Peroxidase are presents in all major body organs, being specially concentrated in liver and erythrocytes.

The antioxydants may well by acting to disminish tissue damage caused by radical reactions induced by toxins in the food supply, excesive amounts of polyunsaturated fatty acids, cosmic radiations, other ionizing radiation, and by exposure to the oxygen in the air, especially if respiratory activity is high. Therefore we analyzed two of that systems: Catalase and Glutathione-Peroxidase systems, seeing that, the fighter / pilots, breathe a mixture very wealthy in oxygen daily. Ours results shown very important increase inboth enzymes over normal ranges.

## CONCLUSSIONS

This work shows the abnormal behaviour of sensitive nervous velocity of conduction in fighter pilots in depending of flight hours. The great comsumption of oxygen could be the reason for that alteration. Wherefore we have begun a research in this way.As a preliminate results, we have found an important increase in Catalase and Glutathione-Peroxidase. According to the early mentioned we affirm that pilot are subjected to great comsumption of oxygen,as shows the increase in protective enzymatic systems in front of oxidations. This hyperoxygenation could be reason for abnormal behaviour of nervous velocity of conduction.

Finally, we suggest the neurophysiology studies for the selection and maintenance of NATO's air crew, because its simpliness, innocuity and precission, as early and / faitful index of Nervous System condition, that it is the most sensible to variations of internal medium.

TABLE IA- Sensitive nervous Velocity of Conduction
Males - Ulnar nerve

| AGE | N.of subjects | Velocity m/sg |
|---|---|---|
| 11-20 | 20 | 59.83±4.89 |
| 21-30 | 37 | 59.97±4.62 |
| 31-40 | 32 | 56.90±3.54 |
| 41-50 | 30 | 55.82±3.97 |
| 51-60 | 31 | 56.36±5.17 |

TABLE IB. Sensitive Nervous Velocity of Conduction
Males - Median nerve

| AGE | N.of subjects | Velocity m/sg |
|---|---|---|
| 11-20 | 20 | 63.15±5.72 |
| 21-30 | 37 | 61.38±4.07 |
| 31-40 | 32 | 58.21±3.94 |
| 41-50 | 30 | 58.10±4.50 |
| 51-60 | 32 | 58.00±6.29 |

TABLE IC. Motor Nervous Velocity of Conduction
Males - Median Nerve

| AGE | N.of subjects | Velocity m/sg |
|---|---|---|
| 11-20 | 18 | 57.92±2.89 |
| 21-30 | 19 | 60.20±6.28 |
| 31-40 | 27 | 57.64±4.65 |
| 41-50 | 28 | 57.13±4.92 |
| 51-60 | 34 | 56.76±5.41 |

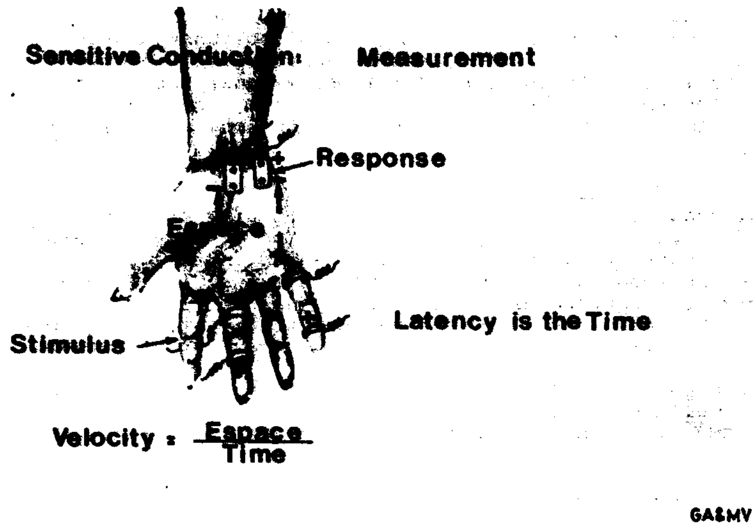



Figure 1 (see text)

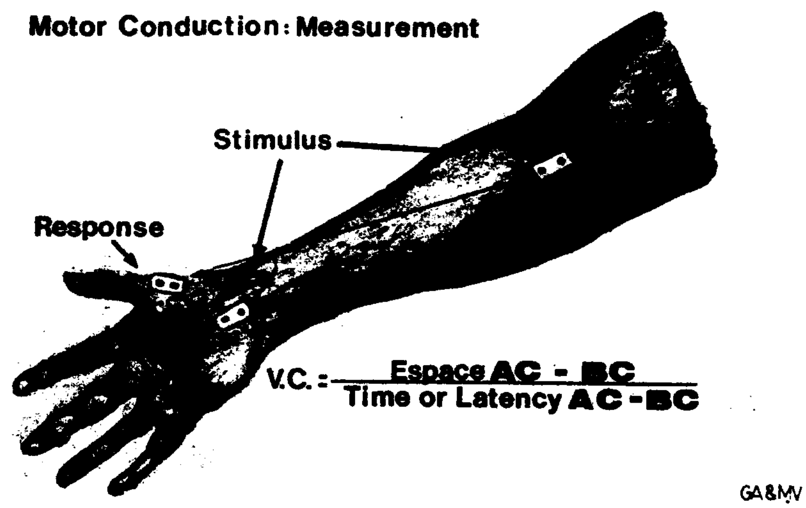



Figure 2 (see text)

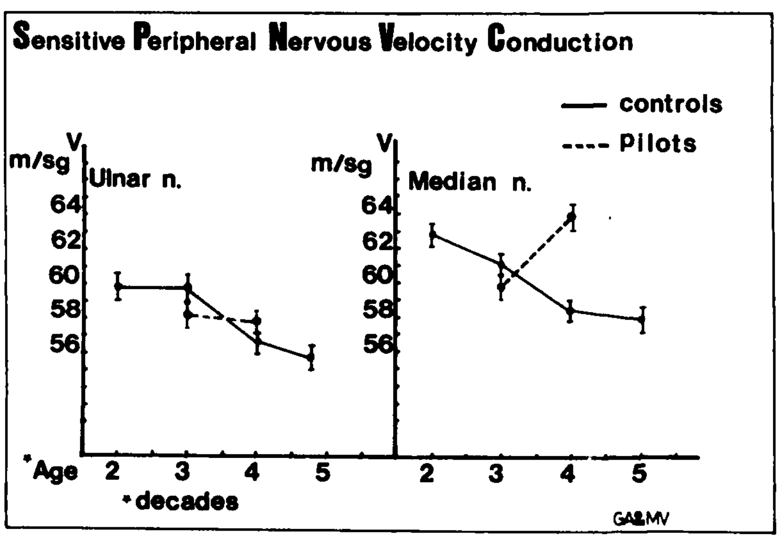



Figure 3

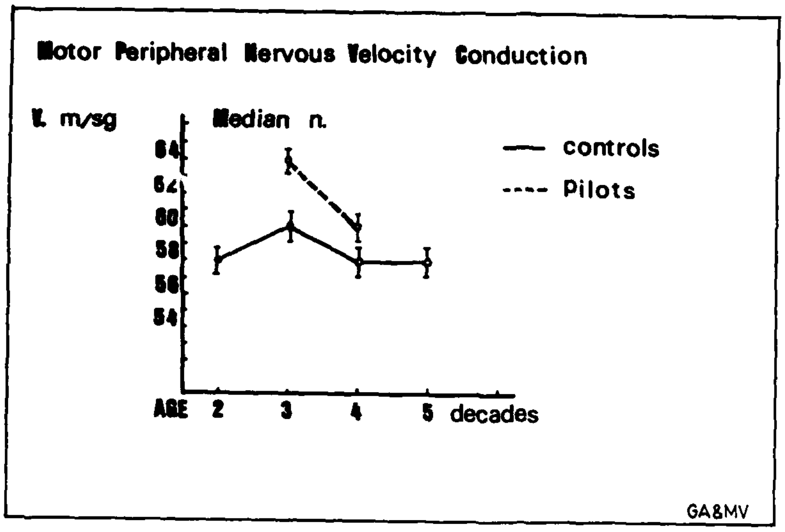



Figure 4

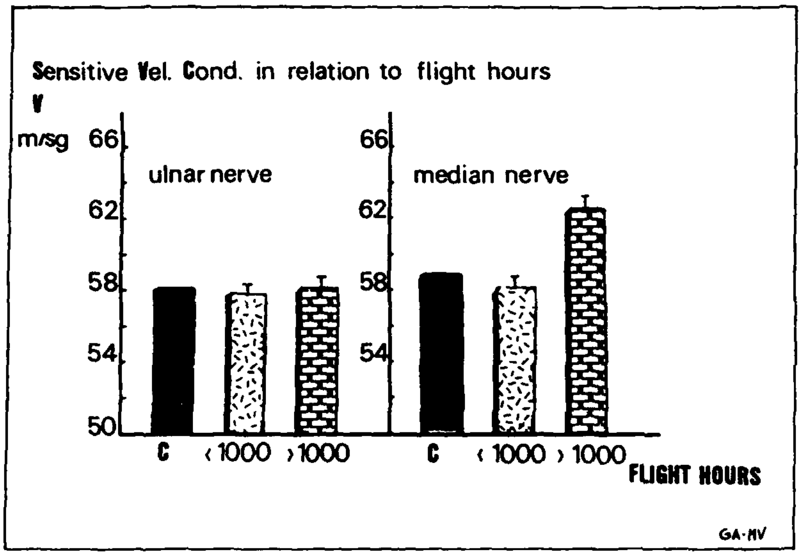



Figure 5

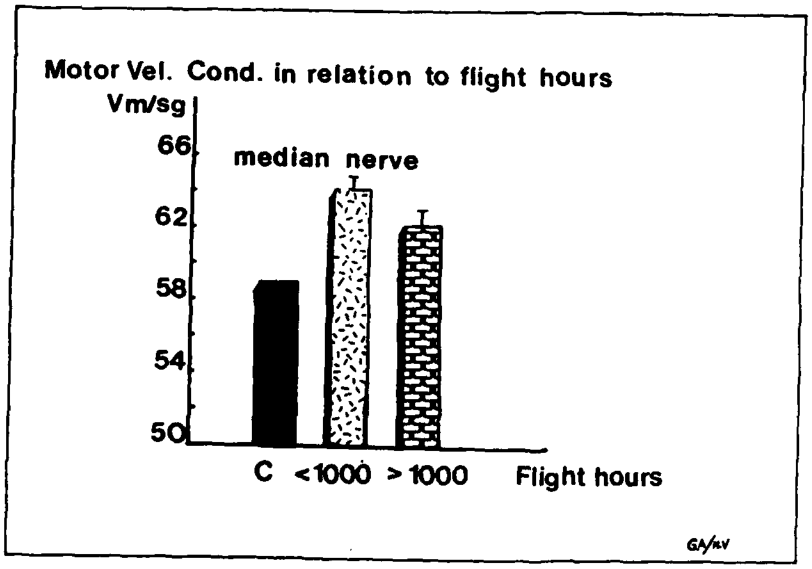



Figure 6

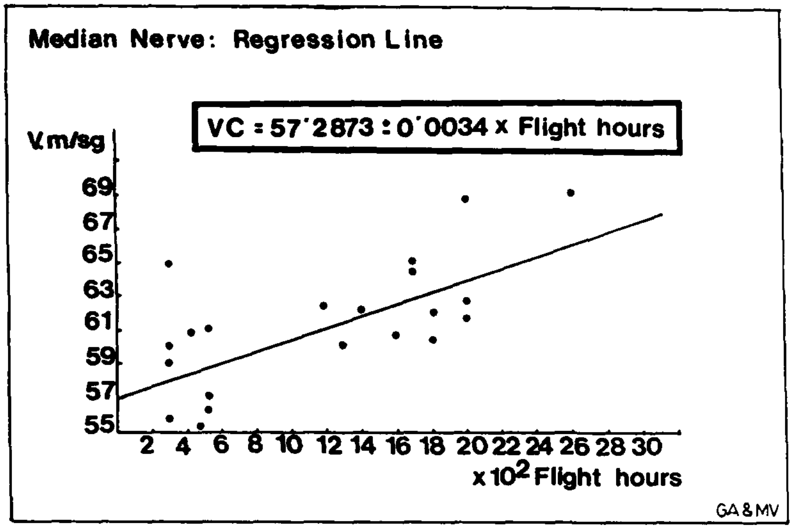



Figure 7

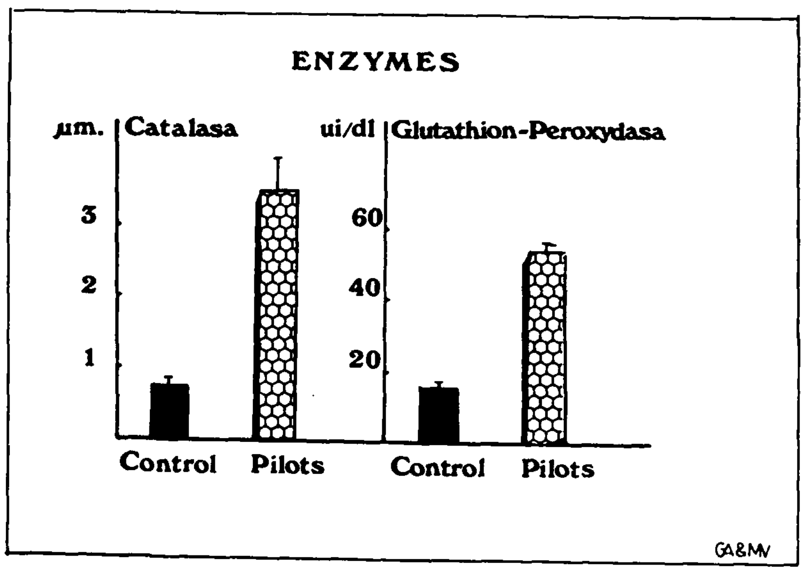



Figure 8

REFERENCES

1.- AEBI,H.:Methods in Enzymology. 1984. Vol 105. (121-126).Academic Press Inc.London
2.- BEHSE, F.;BUCHTHAL, F. and Cols.: "Unmyelinated fibres and Schwam cells of sural nerve in Neuropathy". Brain, 1975, 98 (493-510).
3.- CRUZ MARTINEZ, A; BARRIO, M. and Cols.:"Aspectos electrofisiológicos de la neuropatia urémica. I: Electromiografia, conducción nerviosa motora y sensitiva y respuestas reflejas". Rev.Clin.Esp.,1980.156 (419-424).
4.- FLOHE,L.and GUNZLER, W.A..: Methods in Enzymology.Vol 105.1984.(114-121).Academic Press.Inc. London.
5.- HALLIDAY, A.M.: Evoked Potentials in clinical testing. 1982. Caps 2 and 12.Churchill Livingstone Eds. London.
6.- HALLIWELL,B.;GUTTERIDGE,JMC.: Free Radicals in Biology and Medicine.1985.Caps 2,3 and 8.Oxford University Press. New York.
7.- HARATI,Y.:"Diabetic peripheral neuropathies". Ann. of Int. Med. 107,1987,(546-559)
8.- KRARUP, C.; BUCHTHAL, F.: "Conductions studies in peripheral nerve".Neurobehav. Toxicol.Teratol. (US).7,1985 (319-323)
9.- LOW, P.A.;SCHMELZER, J.D. and Cols: "Experimental chronic hypoxic neuropathy: relevence to diabetic neuropathy". Am.J.Physiol.,250-13,1986 (E94-E99).
10.- LOW, P.A.;TRUCK, R.R. and Cols.: "Prevention of some electrophysiologic and biochemical abnormalities with oxygen supplementation in experimental diabetic neuropathy". Proc.Natl.Acad. Sci. (US),1984 (6894-6898).
11.- LOW,P.A.;WARD, K. and Cols.: "Ischemic conduction failure and energy metabolisme in experimental diabetic neuropathy". Am.J.Physiol. 248-11,1985 (E457-E462).
12.- MORENO VAZQUEZ,J.M.; GARCIA ALCON,J.L. and Cols.: "Estudio de la velocidad de conducción nerviosa periférica enpilotos reactoristas del Ejército del Aire". Med. Mil. de las FAS de España, 44/1,1988 (9-11).
13.- MORENO VAZQUEZ J.M.;GONZALEZ RONCERO,A. and Cols.: "Influencia de la edad y el sexo en la conducción nerviosa periférica". Rev.Esp.Neurologia.1988.
14.- MORENO VAZQUEZ J.M; JIMENEZ HUELVA,J.L. and Cols.:" Neuropatia diabética. I: Influencia del grado de control metabólico en la neuropatia distal. Estudio neurofisiológico. Rev.Esp.Neurologia.(In print).

ACKNOWLEDGMENTS

The authors would like to give special thanks to Dr P. Mena (Human Physiology Department of Extremadura University) for his assistance with enzymatic analysis;Dr M. Molina (Bioestatistics Department of Extremadura University),for his helpful with statistical analysis and all pilots of 23th Wing SAF, for their invaluable assistance.

DISCUSSION

BENSON: You have shown high levels of the enzymes in the pilot population and also higher conduction velocities. Have you correlated individuals' enzyme levels and conduction velocities?

ALCON: No, the pilots were analysed as a group, not individually.

# TRAINING AND SELECTING INDIVIDUALS FOR HIGH LEVELS OF INFORMATION PROCESSING LOAD

**Diane Damos**

*Associate Professor*
*Dept. of Human Factors*
*Inst. of Safety & Systems Mgmt.*
*Univ. of Southern California*
*Los Angeles, CA 90089-0021, U.S.A.*

## SUMMARY

This paper examines techniques for training and selecting pilots to deal with high information processing loads. It begins with a brief review of models of human information processing, then discusses four methods of selecting pilots who can process large amounts of information quickly. Three of these four methods (selection based on the Type A behavioral pattern, measures of specific timesharing abilities, and the specific multiple-task response strategy) are recommended either for immediate use or for more extensive evaluation. Automation, the development of timesharing skills, and the development of flexible visual scan patterns are techniques that could be used to increase a pilot's information processing rate. None of these techniques has a basic research data base sufficient for the development of operational training techniques. All three are, however, promising and should be pursued in a systematic fashion.

In many situations in modern military flying, the amount of information presented to the pilot exceeds his information processing capacity. When the pilot can not process the information in the allotted time, he may decide to ignore some of the information, or he may decide to process the most important information first, leaving the rest until later. Regardless of the strategy the pilot selects, system performance will almost certainly deteriorate.

Currently, system designers must assume that pilots will, at least occasionally, encounter high information processing loads. Modifying or adding hardware to reduce the amount of information to be processed is costly and, in some cases, technologically impossible. The designer must, therefore, find methods of increasing the amount of information the pilot can process in a brief period of time. Basically, the designer has two general methods of increasing the pilot's information processing rate: select pilots who naturally can process large amounts of information quickly and train pilots to process quickly.

The remainder of this paper is divided into a selection section and a training section. The selection section describes existing tests that could be used to select pilots who can processing large amounts of information quickly. The training section describes techniques that could be used to train pilots to process information even more efficiently. Some of the tests and techniques that will be discussed are very new and may need more development. When this is the case, the shortcomings are noted, and topics needing more research are described. Although this paper is divided into two major sections, selection and training are not seen as mutually exclusive. On the contrary, the author believes that pilots should first be selected for their information processing abilities and then trained to use efficient information processing strategies.

The approach taken in this discussion will be guided almost exclusively by data rather than theories of information processing; the limited capacity models (1) and the multiple resources models (2, 3), the most important information processing models from a human factors perspective, have almost nothing to say about performance under high information processing loads. The few statements about processing under high information loads contained in these models can not readily be used to train and select pilots. The types of models proposed by many cognitive psychologists do, however, designate short-term memory as the system restricting the rate of information processing. A technique for circumventing the limitations of short-term memory will be discussed in detail below.

## SELECTING OPERATORS FOR HIGH INFORMATION PROCESSING LOADS

Two major types of tests are available for selection purposes. The first type assesses personality traits and behavioral patterns. The Type A coronary-prone behavioral pattern (4) appears to be the only personality trait or behavioral pattern related to performance under high workload conditions. Type A individuals are characterized as hard-driven, competitive, and achievement-oriented with an extreme sense of time urgency. They are continually involved with achieving more is less time. As a result, Type A individuals are concerned with deadlines, often setting their own deadlines when none are imposed. Individuals who lack these characteristics are referred to as Type Bs. The Type A behavioral pattern can be assessed using two major methods: a structured interview and paper-and-pencil tests. From a practical standpoint, the paper-and-pencil tests should be used for selection purposes; they have reasonably high test-retest reliabilities and typically require less than 20 minutes to complete.

Results from three different lines of research indicate that Type A individuals may be able to deal with high information processing loads better than Type B individuals. First, Type As prefer to work at a more rapid pace than Type Bs (5). This may indicate that Type As naturally prefer a faster rate of information processing than Type Bs. Second, Type A individuals perform better under dual-task conditions (6,7) than Type Bs. Because approximately twice as much information is presented under dual-task conditions as under the comparable single-task conditions, this result again implies that Type As may be able to process more information in a given period of time than Type Bs. Third, Damos and Bloem (7) found that the slope of the function relating correct reaction time to set size in the Sternberg memory search task (8) was twice as large for Type B subjects as for Type As. This result implies that Type A subjects made memory comparisons at twice the rate of Type B subjects, another indication that Type As may processing information more quickly than Type Bs.

Despite the apparent superiority of Type A individuals on information processing skills, they have two shortcomings that may effect their relative ability to process large amounts of information quickly in operational situations. First, they perform no better than Type Bs under time stress (6, 5). Thus, because many tasks in flying are naturally paced, Type As may perform no better than Type Bs. Second, Type As may have more of a tendency to give up in high stress situations than Type Bs if they become convinced that they can not control the stressful event (9).

Several areas of research concerning the Type A behavioral pattern need to be addressed before its use as a selection method can proceed. Few studies have been conducted comparing Type A and Type B individuals on tasks that are required in flying, such as tracking. The basic research data base needs to be expanded appropriately. The tendency to give up in high stress situations has not be observed often, but it clearly should be examined in detail before the Type A behavioral pattern can be used to select pilots for situations that involve high information processing loads.

The second type of selection test assesses skills or abilities. Several different abilities can be involved in the rapid processing of large amounts of information. If the information occurs under multiple-task conditions i.e., the high levels of information are caused by information from more than one source, both general and specific timesharing abilities may be relevant to good performance.

Theoretically, individuals with good general timesharing ability will perform better under multiple-task conditions than individuals with poor timesharing ability regardless of how well they perform on each task alone. During the 1880's and 1890's, a number of distinguished psychologists attempted to isolate a general timesharing ability. Research on this topic continued until the 1920's when interest in identifying this ability ceased, to be revived in the late 1960's. During the last 20 years, several major research efforts have again attempted to isolate a general timesharing ability. Interestingly, despite approximately 100 years of research on this topic and at least a dozen major experiments, little evidence for a general timesharing ability exists. Ackerman, Schneider, and Wickens (10) have discussed many of the methodological problems associated with earlier studies, but the overall impression is still that a general timesharing ability plays a insignificant role in multiple-task performance, if it indeed exists.

Some investigators recently have argued for several specific timesharing abilities rather than one more global ability. The nature of these abilities has never been clearly described although some are assumed to be concerned with certain types of processing, such as parallel versus serial, or with certain kinds of tasks, such as tracking. For example, Braun and Wickens (11) present evidence for five specific abilities (the authors refer to these as components): 1) parallel processing, 2) serial processing, 3) adaptation to rapidly changing control dynamics, 4) performance of different control dynamics, and 5) development of an internal model of a system.

The ability to switch attention between sources of auditory information is another hypothesized specific timesharing ability. A task that measures this ability, the dichotic listening task, has been used in numerous experiments. Some of these experiments have been concerned with basic research issues and have attempted to relate performance on the dichotic listening task to performance on other tasks (11). Others have attempted to use performance on the dichotic listening task to predict performance on complex, operational tasks. Performance on this task has been shown to predict success in flight training (12) and in bus driving (13) although McKenna, Duncan, and Brown (14) have recently questioned the validity of these predictions.

The primary problem with using measures of specific timesharing abilities for selection is that, again, not enough basic research has been performed to determine the validity and reliability of these measures. More applied research is also notably lacking; few studies have attempted to predict performance on complex laboratory tasks or on complex "real-world" tasks based on measures of specific timesharing abilities.
The primary exception to this is, as noted, the research on dichotic listening. The general success of this research is encouraging, supporting the idea that other measures of specific timesharing abilities may be equally useful.

The strategy a person uses to perform two concurrent discrete tasks may also provide a useful selection measure. Three distinct strategies have been identified--a simultaneous, an alternating, and a massed strategy (15). Individuals who adopt a simultaneous response strategy respond to both tasks within some small time period (typically less than 30 ms). Those who adopt an alternating strategy strictly alternate responding to the two tasks. The third strategy, the massed strategy, is more variable, but individuals using this strategy usually make multiple responses to one tasks before responding to the other. The simultaneous and alternating strategies are associated with good multiple-task performance; the massed strategy, with poor performance.

The response strategy may represent a specific timesharing ability related to parallel versus serial processing. More probably, however, it reflects a "style" of information processing under dual-task conditions. Damos, Smist, and Bittner (16) found that subjects could adopt a new response strategy upon request, but in most cases dual-task performance was poorer with the new strategy than with the old one. To date, no measures of personality traits, behavioral patterns, personal history, field dependence/independence, or handedness relate to choice of response strategy. Again, the primary stumbling block to adopting response strategy as a selection test is the lack of both a basic and an applied data base, such as those described above.

## TRAINING OPERATORS FOR HIGH INFORMATION PROCESSING LOADS

Three techniques that could be used to train individuals to process large amounts of information quickly are discussed in this section. The first concerns the development of timesharing skills. Timesharing skills are not, of necessity, related to either a general timesharing ability or specific timesharing abilities. They are assumed to be learned only in multiple-task situations and account for some of the improvement in multiple-task performance that occurs with practice. Little is currently known about these skills. Damos and Wickens (15) identified two timesharing skills, rapid intertask switching and parallel information processing, and found that these skills transferred between different types of task combinations. Jasutis (17) later also found evidence for both the existence and generality of timesharing skills.

Because timesharing skills appear to be generalizable, it should be possible to train them in a laboratory setting and transfer them to an operational setting. Only the two studies discussed above, however, have investigated the training and transfer of timesharing skills. Thus, more research identifying these skills and determining their generalizability to operational situations should be conducted.

The second approach involves the nature of information processing. Currently, information processing is represented on a continuum ranging from controlled processing to automated processing. Controlled processing is believed to be under subject control and is capacity limited. It is slow, effortful, and performed in a serial manner. In contrast, automatic processing in not under subject control and is not limited by the capacity of short-term memory. This type of processing is fast, relatively effortless, and can be performed in parallel with other processes (18). Only tasks that involve consistent processing, i.e., have a constant stimulus-response mapping, can become automatic (19).

The benefits of automated processing are apparent. Automated processing results in faster and more accurate performance than controlled processing. It also bypasses the major information processing bottleneck: short-term memory. The primary drawback in training to automated processing is the extensive practice required. Although performance on any task involving consistent information processing can be automated, few tasks actually have been investigated. Most of these have been laboratory tasks involving visual search (18) although recently several experiments using mental rotation tasks have been conducted (20).

In these experiments subjects have first been given extensive training in two-dimensional rotation of roman letters. They have then performed an identical mental rotation task using either different letters or abstract two-dimensional shapes. Data indicate that the mental rotation of two-dimensional figures becomes automated with practice; i.e., the reaction time is no longer a function of the amount of rotation. In these experiments, subjects often obtain rotation rates of 2000 to 3000 degrees per second. More importantly, these skills in automated rotation appear to transfer although the amount of transfer appears to be related to the similarity of the objects being rotated.

Only one experiment (21) has examined automated processing in an applied context. Subjects were trained to perform a job function common in the telecommunications industry: identifying the geographical region and class of a service problem from a printed document. Subjects received extensive practice in searching a visual field for critical codes presented in specific locations in the field. In the consistent processing condition, specific combinations of letters distributed in a specific spatial relation were always targets. In the variable processing condition, a specific combination in a specific spatial relation might be a target on one trial and a distracter (noise) on the next. Subjects trained under the consistent processing condition detected target patterns faster and more accurately than subjects trained under the variable processing condition. Additionally, these subjects were much less affected by the number of targets they were searching for (one, two, or three) than those in the variable processing condition.

Unlike many of the tests discussed in the selection section and some of the techniques discussed in this section, automation is well understood theoretically and ample laboratory data on visual tasks exist. The benefits of automation are also clear: Performance becomes fast, accurate, and is not affected by the processing load inherent in the task. The existing data base has three shortcomings that limit the immediate use of automation techniques to increase the information processing rate of pilots. First, few data exist on many of the tasks that are of interest in an aviation context. Many of these tasks do not have consistent stimulus-response mappings and no amount of practice may result in automation. Second, automated tasks need to be studied in a multiple-task context. Theoretically, an individual should be able to perform an automated task with another task without a performance decrement on either task. To date, automated visual search tasks have only been performed with other visual search tasks (22). Indeed, no decrement has occurred on either tasks, but this result may not be typical of other task combinations. Third, more applied experiments showing the usefulness of this concept in operational settings are necessary.

The third training technique involves the pilot's scan pattern. Novice military pilots usually receive instruction in efficient scan patterns as part of undergraduate pilot training. In U.S. Air Force Undergraduate Pilot Training, for example, the instructor pilot demonstrates an efficient scan pattern for each instrument maneuver taught. Although an efficient pattern is demonstrated, the instructors emphasize adaptability and encourage the student to experiment with different cross checks to find the student's optimal scan pattern.

The implication of this training is that experienced pilots, such as military instructors, should have efficient, consistent scan patterns that are idiosyncratic. Interestingly, DeMaio, Parkinson, Leshowitz, Crosby, and Thorpe (23) found that, when confronted with a novel situation, instructor pilots detected deviations more quickly and accurately than student pilots but showed no evidence of a consistent scan pattern. DeMaio, Parkinson, and Crosby (24) found that military flight instructors do not appear to process information from aircraft displays sequentially. Rather, they appear to perform some of the processing in parallel. The results from these two experiments may indicate that, with extensive practice, pilots can obtain information from displays automatically. This hypothesis is consistent with the results discussed earlier on automatic processing; almost all demonstrations of automatic processing have occurred on visual search tasks. Thus, it may be possible to train pilots' visual "scanning" behavior to automaticity using techniques that have already been developed.

## CONCLUSION

This paper discusses methods to select and train pilots to process large amounts of information quickly. The most promising selection tests assess the Type A behavioral pattern, specific timesharing skills, and the response strategy used to perform concurrent tasks. Both the Type A pattern and the response strategy could be operationally evaluated immediately; both of these measures have reasonable basic research data bases although more applied research on both topics would be desirable. Measures of one specific timesharing skill, attention switching, have already been used to select pilots and professional drivers. More basic research is need, however, to identify and characterize other specific timesharing abilities. Selection based on a measure of general timesharing ability is not feasible at this time. No reliable measure of a general timesharing ability currently exists, and the data base gives little hope that such a measure will be forthcoming.

Three techniques for training pilots to deal with high information processing loads were discussed. One of these, the development of efficient scan patterns, has been taught for approximately 40 years in military aviation. Recent findings suggest, however, that more efficient scan patterns could be developed quickly by training to automaticity. The other two techniques, development of timesharing skills and automatic processing, need to be systematically investigated from a training perspective. Training techniques for aviation applications could oe available shortly if such a systematic investigation were made.

## REFERENCES

1. Broadbent, D. *Perception and Communications.* London: Pergamon, 1958.
2. Kahneman, D. *Attention and Effort.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.
3. Wickens, C. *Engineering Psychology and Human Performance.* Columbus, OH: Charles E. Merrill, 1984, 301-305.
4. Freidman, M., and Rosenman, R. *Type A behavior and your heart.* New York: Knopf, 1974.
5. Frankenhaeuser, M., Lundberg, U., and Forsman, L. Dissociation between sympathetic-adrenal and pituitary-adrenal responses to an achievement situation characterized by high controllability: Comparison between Type A and Type B males and females. *Biological Psychology, 10,* 1980, 79-91.
6. Damos, D. The relation between the Type A behavior pattern, pacing, and subjective workload under single- and dual-task conditions. *Human Factors, 27,* 1985, 675-680.
7. Damos, D.,and Bloem, K. Type A behavior pattern, multiple-task performance, and subjective estimation of mental workload. *Bulletin of the Psychonomic Society, 23,* 1985, 53-56.
8. Sternberg, S. High-speed scanning in human memory. *Science, 18,* 1964, 953-956.
9. Glass, D., Snyder, M., and Hollis, J. Time urgency and the Type A coronary-prone behavior pattern. *Journal of Applied Social Psychology, 4,* 1974, 125-140.
10. Ackerman, P., Schneider, W., and Wickens, C. Deciding the existence of a time-sharing ability: A combined methodological and theoretical approach. *Human Factors, 26,* 1984, 71-82.
11. Braun, R., and Wickens, C. Time-sharing revisited: Test of a componential model of the assessment of individual differences. *Ergonomics, 29,* 1986, 1399-1414.
12. Gopher, D., and Kahneman, D. Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skills, 33,* 1971, 1335-1342.
13. Kahneman, D., Ben-Ishai, R., and Lotan, M. Relation of a test of attention to road accidents. *Journal of Applied Psychology, 58,* 1973, 113-115.
14. McKenna, F., Duncan, J., and Brown, I. Cognitive abilities and safety on the road: a re-examination of individual differences in dichotic listening and search for embedded figures. *Ergonomics, 29,* 1986, 649-663.
15. Damos, D., and Wickens, C. The identification and transfer of timesharing skills. *Acta Psychologica, 46,* 1980, 15-39.
16. Damos, D., Smist, T., and Bittner, A. Individual differences in multiple-task performance as a function of response strategy. *Human Factors, 25,* 1983, 215-226.
17. Jasutis, J. Department of Psychology, Purdue University. *A comparison of transfer of specific task skills and timesharing skills.* 1978. Unpublished master's thesis.
18. Schneider, W., and Fisk, D. Automatic category search and its transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 1984, 1-15.
19. Fisk, D., and Schneider, W. Consistent attending versus consistent responding in visual search: Task versus component consistency in automatic processing development. *Bulletin of the Psychonomic Society, 22,* 1984, 330-332.
20. Grose, E., and Damos, D. Automaticity and the transfer of mental rotation skills. *Proceedings of the Human Factors Society 32nd Annual Meeting.* Santa Monica, CA: Human Factors Society, in press
21. Myers, G., and Fisk, A. Training consistent task components: Application of automatic and controlled processing theory to industrial task training. *Human Factors, 29,* 1987, 255-268.
22. Schneider, W., and Fisk, A. University of Illinois, Human Attention Research Laboratory. *Dual task automatic and controlled processing in visual search, can it be done without cost?* 1980. Technical Report No. 8002.
23. DeMaio, J., Parkinson, S., Leshowitz, B., Crosby, J., and Thorpe, J. Air Force Human Resources Laboratory. *Visual scanning: comparisons between student and instructor pilots.* 1976. Technical Report No. AFHRL-TR-76-10.
24. DeMaio, J., Parkinson, S., and Crosby, J. A reaction time analysis of instrument scanning. *Human Factors, 20,* 1978, 467-471.

## ACKNOWLEDGMENTS

The preparation of this manuscript was partially supported by Contract No. N00014-86-K-0119 from the Naval Medical Research and Development Command. Dr. Guy Banta was the contract monitor.

DISCUSSION

MAAT: I would like to ask a question about specific time sharing ability, the third point you mentioned with respect to selection measures. How specific are your specific time sharing abilities? I mean, could you apply the same specific time sharing abilities to operators in process industries?

DAMOS: We believe so. The reason why we believe that is fundamental to psychology. You believe that certain human beings have certain ways of processing information. The pilot represents a very skilled person for certain types of information processing. He may be naturally faster for certain types of information processing, but we believe that the building blocks are the same. If you can identify them in one place you should be able to use them somewhere else. So far that seems to have worked.

PSIMENOS: Do you have any observations in real groups of pilots regarding this trait? Have you compared the scores of pilots who do well with those that fail. What is the application of this test to predicting the real performance of pilots?

DAMOS: Most of you are aware that the United States Air Force and the United States Navy are beginning to look at computerised tests. Those tests are primarily tests of information processing. So they are looking at many of these fundamental building blocks, and are busy gathering data on pass/fail for people who have higher versus people who have lower scores. Now we do have one technical report out that I did with Lt Gibb about two years ago. This compared performance on a number of these various tasks of very highly skilled, aggressor fighter pilots with that of student pilots who had not yet started flight training. We did not see a lot of difference, but we had two problems. One was that the pilots were older and, therefore, had some natural slowing of reaction times, which confounded the results. Where we did see differences between the two groups, these were obviously skill related and we could not be dissembled at the time. For example, the tracking task performance of the very highly skilled pilots, of course, looked different from that of the student pilots.

PSIMENOS: Have any observations been made between a group that failed in training and another group that succeeded?

DAMOS: They are just getting that data now. As a matter of fact, I think Lt Dolgin will have an answer.

SIEM: We have just recently conducted a study in which we compared undergraduate pilots who after a year of training were recommended for fighters with another group of undergraduate pilots who had passed training but were not recommended for fighters. The fighter-recommended pilots manifested a superior performance in the mental rotation test, and in speed and accuracy on both the Sternberg task and a time sharing task. On all three cognitive tests the fighter-recommended pilots were significantly different in terms of their performance. These tests would appear to be applicable in selection.

THACKRAY: You mentioned the Type A personality as preferring a high rate of information load; this is what our Symposium is really all about. However, in our work with monitoring tasks, primarily in the context of increasing automation, what we find is that there is no difference between type A and type B individuals. Perhaps this is because we select them on their ability to perform these rather passive tasks. I do not know if you found this or not?

DAMOS: No, I was looking at, not monitoring, information processing. I was not aware of your data.

HULME: With the advent of automated systems in 10 years' time, or whatever, do you think you are going to have to change your selection procedures and the requirements for personality types?

DAMOS: Well, we do not select now based on personality types.

HULME: Selection procedures then.

DAMOS: I think we are seeing a move towards it. The move is away from paper and pencil tests towards information processing measures; this move is just beginning but it is a major change.

MAAT: I would like to ask you a question about training of automatic processing skills. Do you think it should be diffferent from training regular things or general education? In your opinion what are the differences between learning by instruction from teachers and learning from training devices?

DAMOS: Automatic processing must be trained in a very different manner, and the training that has been successful is all computer based. It depends on the extreme repetition of a stimulus so that a direct connection between the stimulus and the response is established. We think that when the connection is sufficiently strong short term memory is bypassed. I have never seen any training for automatic processing done using normal instruction. The type of pairing you need and the amount of repetition almost precludes that. It does take time, but you do not have to do it in real time. Training on air interceptor tasks, using automated processing methods, with non-real time presentation of problems, has worked very well. In fact, that is the most successful application of it to date.

# PERSONALITY ASSESSMENT IN AVIATION SELECTION

Daniel L. Dolgin, Ph.D., LT, MSC, USNR
Naval Aerospace Medical Research Laboratory
Pensacola, Florida USA 32508-5700

## SUMMARY

A comprehensive review of the personality literature as it relates to aircrew selection was conducted. The purpose of the study was to identify specific tests that warrant further research as potential prediction instruments. Aviation selection techniques in the U.S. Navy and U.S. Air Force were reviewed. Individual tests used in aviation selection are discussed in terms of their results. The advent of performance-based personality assessment in the 1970s is examined, and implications for future test development are explored. The majority of personality instruments reviewed were invalid for pilot selection. In some cases, methodological difficulties may have obviated more promising results. Recommendations are made for continued research with several tests that appear to be both effective in pilot selection and psychometrically sound. Those recommended selection tests include the Defense Mechanism Test because of its effectiveness in predicting pilot training success and safety in the Swedish and Danish forces. The Personality Research Form is recommended due to both its psychometric construction and current research efforts that are ongoing in the Canadian Armed Forces and U. S. Air Force. The Locus of Control is also proposed for both closer and continued attention. Other recommended selection instruments include the Work and Family Orientation Questionnaire and Extended Personality Attributes Questionnaire. Safety in aviation is also addressed as a major, emerging area of interest in the 1980s.

## WHY THE FAILURES: METHODOLOGICAL PROBLEMS AND ISSUES

The 1980s has reflected renewed interest in personality as it relates to aviation selection. Aptitude testing alone cannot predict all of the flight failures, thus the exploration of personality variables and decision making styles to improve the selection of military aviators becomes more critical. The purpose of this presentation is to examine some of the methodological problems and issues that have perhaps affected the results of personality assessment in aviation.

### Broad Criterion Measures

Most efforts to increase the predictive validity of aviation screening systems have some inherent methodological problems. Typically, test measurement variables are related to global criterion performance measures in training such as graduation/elimination or composite flight grades. Such performance criteria, although highly useful, have several undesirable psychometric properties and may obscure the components of skilled performance or behavioral attributes associated with the selection test measure. Presumably, a given test measure may be highly predictive of a critical performance dimension during some phase or component of flight training, but the insensitivity or impracticality of the performance criterion may yield low correlations and a consequent dismissal of the test's predictive power. Helmreich et al. (1) further point out that different combinations of predictors relate to quite different measures of performance at different points in time.

Previous efforts to investigate the use of personality indices characteristically have been piecemeal and have examined only one or a few tests related to a given overall flight performance criterion, usually a composite measure at the conclusion of initial flight training. Additionally, the vast majority of investigations used subjects that already were preselected on standard selection measures. Thus, in many cases, only simple relationships between a personality measure and a singular criterion are presented. Relatively few studies contain multiple regression models of the initial candidate selection variables. It cannot be determined whether a particular personality variable actually adds unique variance to prediction of training success beyond the initial selection measures. Unfortunately, efforts to relate specific predictors to reliable subcomponents of overall flight grades in primary training proved unsuccessful (2,3). The authors concluded that reliable clusters of performance criteria were embedded in the overall cumulative flight grade. This was attributed to a wide disagreement among instructor pilots as to which individual measures of flight performance were used most in evaluating differences in student performance.

Earlier research to develop subcriteria embedded in the more global criterion of graduation/elimination met with similar failure. The Army Air Forces Pilot Project (4) attempted to develop subcriteria against which specific selection measures of aptitude could be validated, but problems of restricted range in grading flight performance were identified as a major reason for the lack of success (2). Flight students were evaluated purely in subjective terms in one of several categories, A-F, with the majority of grades assigned a 'C.' This was due to the emphasis on determining which students would not successfully complete flight training as opposed to providing a normal distribution of grades to differentiate among students who were successful.

Related to the restricted range problem in the flight criteria is the halo effect phenomenon (4). Typically, check pilots and instructor pilots access and utilize information about a student's past performance in arriving at a current evaluation. Correlations between performance measures for different maneuvers and procedures tend to be high, suggesting the presence of a strong halo effect. Grading tendencies of flight instructors to the average or "norm" can also be reflective of a de-emphasis towards comparing successful students during primary training. Current military primary flight training systems also require instructor pilots to provide a written explanation when they assign a grade other than average (i.e., below average, above average, and unsatisfactory). In short, evaluating any maneuver or performance as other than average requires additional time-comsuming documentation by the instructor. A related issue is the reliability of assigned flight grades. The importance of this methodological concern to pilot selection was noted over 40 years ago (5). Studies conducted during the Army Air Force Pilot Project indicated that landing performance measures correlated near zero for repeated measurements on the same maneuvers during different days using different aircraft and different instructors with the same students.

Homogeneous Population

A methodological problem of validating personality instruments also exists with the candidate population itself. Most personality inventories and clinical diagnostic tools were developed for testing heterogeneous groups. Military aviation candidate populations tend to be comparatively homogeneous. Typical entrance requirements include a 4-year college degree, rigid medical requirements, and initial aviation screening tests. Application to a flight training program in itself reflects a general interest in aviation. Most applicants are males in their early twenties. Military age standards partially account for the similar age factor found in the candidate population. All of these factors combine to result in a rather unique, homogenous population that severely restricts the sample population at the outset.

Honeymoon Effect

Others maintain that personality measures can only effectively predict actual job performance and not training performance. Helmreich (6) emphasizes that "deficiencies in the criterion lead to overemphasis on some predictors and the neglect of others." Helmreich et al. (1) reported a link between personality and performance, which they called the "honeymoon effect" of motivation on performance. They believed that the honeymoon effect was the maximum effort that many job prospects exert in order to obtain a coveted position or job. The underlying personality dispositions become significant determinants of behavior only after the "honeymoon" period ends. Their study demonstrated a major weakness in using initial training performance as the selection criterion. In the same study, personality and motivational factors measured prior to employment proved to be good predictors of job performance. This prediction was obtained only after the subjects had been out of training and on the job for more than 3 months. The predictors were unrelated to performance in training and after initial release to the workforce.

Helmreich (6) administered the Extended Personal Attributes Questionnaire (EPAQ), developed by Spence et al. (7), and the Work and Family Orientation Questionnaires (WOFO), developed by Helmreich and Spence (8), to a group of pilots. The EPAQ measures positive and negative clusters of instrumental and expressive traits; the WOFO evaluates three aspects of achievement motivation and interpersonal competitiveness. These personality measures were compared to ratings by check airmen (pilots who evaluate actual and simulated flight performance of other pilots). The results indicated that the trait constellations of instrumentality and expressiveness, along with components of achievement motivation, were significantly related to this operational criterion. The better pilots scored higher on instrumentality, expressivity, and high mastery needs, while poorer pilots scored higher on aggressiveness.

Response Bias

Test response bias is often cited as responsible for the lack of validity in predicting a flight training criterion (Anastasi, 9). Test response bias is the inability to obtain a true measure of an individual's character, which is usually attributed to response sets. Social desirability or "faking" is the response set or attitude that has received the greatest attention. As Anastasi (9) pointed out, respondents can easily detect the most socially desirable or acceptable response choices in the majority of personality inventories. In military aviation testing scenarios, candidates generally will respond to create the most impressive image of themselves or to their perception of the "aviator personality." These circumstances provide very little variance on personality measures between respondents (North and Griffin, 10). Such acquiescence, or the tendency to respond in a consistent but inaccurate fashion, is an additional response set that can affect the predictive validity of an instrument. Many personality inventories are structured such that all "true," "yes," "a," et cetera, responses are keyed positively for the personality dimension of interest. This type of format is susceptible to some respondents answering "yes," "no," or "middle-of-the-road" for all questionnaire items. This type of response pattern does not accurately reflect the trait being assessed.

Commercial Availability

A final consideration, which is often neglected in the use of personality instruments for personnel selection, is commercial availability. Assuming that a particular personality test does meet aforementioned criteria, the predictive value of the instrument will probably decline steadily within a short period of time, as is common with all measurement devices. Nonetheless, the commercial availability of personality instruments compromises test validity and provides an impetus for accelerated deterioration. This is especially true in circumstances where the "score" on the instrument may determine acceptance or rejection into a military flight training program. The fact that job candidates fake personality inventories to gain employment is well established (Green, 11; Stricker, 12). Within 2 years, preparation and "coaching" for the instrument may be found in commercially available guides (i.e., Officer Candidate Tests, Wiener, 13), and the test could be compromised. Considering these disadvantages, we recommend investigations of non-inventory techniques and methods of measuring personality that might provide useful additional predictions to measures of aptitude. One such approach could be toward the development of measures in which the personality dimension of interest is "masked" or concealed from the candidate.

## EMERGENCE OF AUTOMATED BEHAVIOR-BASED INVENTORIES

The need to improve the selection of military aviation applicants, along with recent advances and innovations in computer technology and psychological theory/measurement (Anastasi, 9), have combined to stimulate interest in computerized assessment. This new emphasis is partly responsible for the use of performance tasks, rather than paper-and-pencil tests, to avoid verbal and cultural biases. In the past decade, several computer-based experimental aviation selection test batteries have evolved, along with an interest in reaction time and response time measures as dependent variables.

U.S. Air Force Basic Attributes Tests (BAT)

In 1981, the United States Air Force began a large-scale effort, known as the Basic Attributes Tests system or "BAT," to determine the validity of a computer-based test battery for pilot selection and classification (Kantor and Bordelon, 14). The BAT battery originally consisted of 15 component tests. Although the primary emphasis of the BAT was directed toward measuring psychomotor, cognitive, and perceptual skills, six tests were included to measure personality and attitudinal characteristics. Personality tests that were included or developed were the Dot Estimation Task, Risk-taking, Embedded Figures, Self-crediting Word Knowledge, Activities Interest Inventory, and Automated Aircrew Personality Profiler.

Recently, Siem et al. (15) evaluated five of the personality instruments of the BAT. Data on the Automated Aircrew Personality Profiler were not available. The personality tests were administered to 883 Air Force pilot candidates to assess their utility in predicting training outcome (pass/fail) and advanced training recommendation (fighter or non-fighter aircraft). Both criteria were treated as dichotomous variables. Acceptable reliabilities were reported for all five measures for use as selection instruments.

No single test or individual dependent measure displayed a consistent pattern of validity to both criterion measures. The test for self-confidence (Self-crediting Word Knowledge) appeared to be the only instrument that contributed to predicting successful completion of flight training, with successful candidates demonstrating more caution. The only dependent measure that even exceeded a correlation of .10 with the pass/fail criterion was the correct response reaction time for the Self-crediting Word Knowledge task ($r$ = .12, $p$ < .001). The multiple correlation for the Self-crediting Word Knowledge test was .14. No measure displayed a significant relation to instructor pilot recommendation. Although significant differences were not observed, data comparing 259 attrites with 488 successful graduates indicated a general trend toward cautious responding by students who completed training. These candidates chose fewer high-risk items on the Activities Interest Inventory, required more time, and completed fewer trials on the Dot Estimation Test, and they had higher percentage-correct scores for the Dot Estimation Test. The investigators interpreted these findings, taken in conjunction with the results of the Self-crediting Word Knowledge task, as a more cautious decision-making style on the part of successful candidates. Their interpretation, however, is not supported by results from the Risk-taking Task, which was intended to measure risk tendencies in decision-making.

In summary, the analyses of personality variables under investigation by the Air Force show very little promise for use in selecting or classifying aviation candidates. Further work is ongoing at the Air Force Human Resources Laboratory in San Antonio, Texas, to determine if the Self-crediting Word Knowledge Task adds unique variance to the current prediction model, even though only a weak relationship exists between the instrument and the graduation/elimination criteria. Additional research efforts are focused on improving the existing Self-crediting Word Knowledge Test and evaluating the test's construct validity. To assess specifically what the test is measuring, more traditional personality tests of characteristics such as self-confidence (Spence et al., 7) are being administered to Air Force flight personnel who have varying levels of experience.

U.S. Navy Performance-based Personality Tests

Shull et al. (16) conducted an initial validation of the Navy test for measuring risk-taking tendencies in 440 student naval aviators. The Navy risk test is essentially a computer-based gambling task consisting of 3 sessions with 10 trials in each session. For each trial, the subject is presented with a matrix of squares identified by numbers. At the beginning of each trial, one square is a penalty square, which causes a loss of points, and nine are reward squares. During session 2, two randomly selected penalty squares (for each trial) provide an opportunity to assess changes in response strategy to a more "risky" situation. The subject is allowed to select any of the squares, one at a time, and if the selected squares contain a payoff (points), the subject may keep them. Measures indicating increased risk-taking consist of increases in number of responses made (squares selected) and decreased response latency in making those selections. Results from the risk test were compared to students' raw scores on the Navy's primary flight candidate selection battery and actual grades from flight training. The number of squares selected during the first session and the graduation/elimination criteria were significantly correlated, which indicated that increased risk-taking may be associated with completing primary flight training. The authors also found moderately significant correlations between this particular measure and both the aviation indoctrination and cumulative flight grade scores, although in a direction indicating that decreased risk-taking is associated with higher grades in these areas. If present results are any indication, this test or some revised version of it may hold promise as an effective pilot candidate screening device. A U.S. Air Force study (Siem et al., 15), however, found no relationship between risk taking behavior and pass/fail outcome with a sample of 883 pilot candidates.

RECOMMENDED PERSONALITY TESTS

A primary purpose for the current review of personality research was to identify specific tests that warrant further research as potential prediction instruments. Based on the review of past and present instruments utilized in the selection of pilots, the following tests are recommended for continued research. These tests appear to be both effective in pilot selection and psychometrically sound.

Those recommended tests include the Defense Mechanism Test (DMT) because of this instrument's effectiveness in predicting pilot training success and safety in the Swedish and Danish forces (17). The DMT is a projective personality test that has been used operationally in Scandanavian countries for the past decade. The concept of the DMT in the prediction of success in flight training is that the use of certain defense mechanisms may limit the amount of "psychological" energy available for handling external stress. Because the military flight training environment is highly stressful, a flight candidate with intense defenses might not immediately recognize a dangerous situation. Although the DMT is designed for individual administration and requires 1.5 to 2 h testing time, previous success with the instrument warrants further study. The Personality Research Form (18) is recommended due to its psychometric construction (19) and promising results in the Canadian Armed Forces (20,21) and the U.S. Air Force (22,23). The Cattell 16PF (24) has been used successfully (25-27) to predict success in flight training. Lester and Bombaci (28) found a significant relationship between "hazardous thought patterns" and 16PF scores. As a result of these studies, the 16PF stands out as a personality instrument requiring further investigation. Another test that has achieved some success is the Locus of Control (29). The Locus of Control is a brief questionnaire of 23 items and is easily automated for computer administration. Results (30) indicate that pilots are significantly more internally controlled than the general U.S. population. Developed by Spence et al. (8), the Work and Family Orientation Questionnaire (WOFO) has been related successfully to pilot performance (1). The WOFO operationalizes achievement motivation into components of mastery needs, desire to undertake new and demanding tasks, work orientation, satisfaction with hard work and task completion, competitiveness, and concern with outperforming others in interpersonal situations. A final recommended instrument is the Extended Personality Attributes Questionnaire (EPAQ: 6,7). The EPAQ has typically been employed in research concurrently with the WOFO.

THE ROLE OF PERSONALITY IN AVIATOR SAFETY

Typically, personality research in aviator safety has been directed toward identifying the "accident-prone" aviator. "Accident proneness" is not a stable characteristic and is situationally based (Alkov et al., 31; McGuire, 32). Measurement of the tendency to be accident liable or susceptible would thus be difficult because the tendency varies with time. Increased risk-taking tendencies that result in mishaps would only emerge as a result of situational circumstances in conjunction with an inability to cope with increased stress levels. Alkov et al. (31) suggest that inadequate techniques for coping with stress, rather than cumulative life stress, account for the increased levels of accident susceptibility. They (31) compared pilots who were causally involved in mishaps to aviators involved in mishaps with no culpability. They found that pilots who made errors resulting in mishaps were poorer leaders, were less mature and stable, had undergone a recent personality change, and were experiencing problems with interpersonal relationships. The authors concluded that aircraft mishaps may be attributable to the non-introspective personality, but the data are post-hoc and are not based on a prediction model. Aviators involved in aircraft accidents were interviewed by accident investigation board members, superiors, peers, and family. Information provided by the respondents was colored by the aviator having been involved in a mishap. Using personality devices to predict which individuals would be involved in future aircraft accidents would be difficult and require enormous sample sizes due to the relatively low incidence of mishaps.

Jensen and Benel (33) reviewed the literature pertinent to aviation accident data from 1970 through 1974. Their conclusions were 1) erroneous pilot decision-making was a factor in 35% of all non-fatal aviation accidents, and 2) faulty decision-making played a definite role in 52% of fatal mishaps. The authors noted that research on pilot judgment was sparse with little if any systematic work having been conducted. They maintain that pilot judgment is trainable and can be objectively evaluated. In conclusion, Jensen and Benel speculate that faulty judgment might result from a pilot's proclivity to situational influences such as peer reactions, fear of failure, and censure from superiors or family members.

Lester and Bombaci (28) examined the construct validity of five "hazardous thought patterns" hypothesized to mediate pilot judgment. The hazardous thought pattern concept is the result of an investigation carried out by the Federal Aviation Administration and Embry-Riddle Aeronautical University (ERAU). In response to the Jensen and Benel (33) study, ERAU investigators sought to isolate the specific thought patterns that might serve as the precursors to faulty pilot judgment using a literature review and consultation with experts in the behavioral sciences. Five hazardous thought patterns were identified as anti-authority, impulsivity, invulnerability, macho, and external control or resignation. A 10-item self-assessment inventory was designed to assess the hazardous thought patterns concept. Evaluating a sample of 35 civilian pilots, Lester and Bombaci (28) observed a significant relationship between hazardous thought patterns and scores on both the Cattell 16PF integration/self-concept control scale and the Rotter LOC scale. Lester and Bombaci recommended that additional research examine the way in which situational influences interact with pilot personality.

## CONCLUSIONS

The development and application of personality tests have unique opportunities, as well as special difficulties, that might not be encountered with aptitude testing. For example, test faking and malingering are problems specific to the personality assessment domain. A related problem is the greater situational specificity pertinent to personality test performance. Behavior measured by personality instruments is also more susceptible to change over time than is behavior measured with aptitude tests. Attempts to improve personality assessment have included computerization, the development of verification and correction scales, keying certain items against specific criteria, masking the dimension of interest, and the application of factor analysis as a way of isolating more specific trait categories. The future of personality testing in aviation selection appears most promising in the domain of computer administration and in concealing the personality trait of interest.

## REFERENCES

1. Helmreich, R.L., Sawin, L.L., and Carsrud, A.L. The Honeymoon Effect in Job Performance: Temporal Increases in the Predictive Power of Achievement Motivation. Journal of Applied Psychology, 71, 185-188, 1986.

2. Dolgin, D.L., Gibb, G.D., Nontasak, T., and Helm, W.R. Instructor pilot evaluations of key naval primary flight training criteria. NAMRL-1331, Naval Aerospace Medical Research Laboratory, Pensacola, FL, June 1987.

3. Shannon, R.H. and Waag, W.L. Toward the development of a criterion for fleet effectiveness in the F-4 fighter community. NAMRL-1173, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1972.

4. Melton, A.W. Apparatus Tests: Army Air Forces Aviation Psychology Program Research Report No. 4. U.S. Government Printing Office, Washington, DC, 1947.

5. Miller, N.E. Psychological Research on Pilot Training: Army Air Forces Aviation Psychology Program Research Report No. 8. U.S. Government Printing Office, Washington, DC, 1947.

6. Helmreich, R.L. Pilot selection and training. Paper presented at the American Psychological Association Annual Meeting, Washington, DC, 1982.

7. Spence, J.T., Helmreich, R.L., and Holahan, C.K. Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. Journal of personality and Social Psychology, 37, 1673-1682, 1979.

8. Helmreich, R.L. and Spence, J.T. "The Work and Family Orientation Questionnaire: An Objective Instrument to Assess Components of Achievement Motivation and Attitudes toward Family and Career. JSAS: Catalog of Selected Documents in Psychology, 8, MS 1677, 1978.

9. Anastasi, A. Psychological Testing (5th ed.). Macmillan, New York, NY, 1982.

10. North, R.A. and Griffin, G.R. Aviator selection 1919-1977. NAMRL SR 77-2, Naval Aerospace Medical Research Laboratory, Pensacola, FL, October 1977.

11. Green, R.F. Does a selection situation induce testees to bias their answers on interest and temperment tests. Educational and Psychological Measurement, 1951, 503-515, 1951.

12. Stricker, L.J. Test-wiseness on personality scales. Journal of Applied Psychology Monograph 53, 1-18, 1969.

13. Wiener, S. Officer Candidate Tests. Arco Publishing Co., New York, NY, 1985.

14. Kantor, J. and Bordelon, V. The USAF pilot selection and classification research program. Aviation, Space, and Environmental Medicine, 56, 254-257, 1985.

15. Siem, F.M., Caretta, T.R., and Mercatante, T.A. Personality, attitudes and pilot training performance. AFHRL TP-87-36, Air Force Human Resources Laboratory, San Antonio, TX, 1987.

16. Shull, R.N., Dolgin, D.L., and Gibb, G.D. The relationship between flight training performance, a risk assessment test, and the Jenkins Activity Survey. NAMRL-1339, Naval Aerospace Medical Research Laboratory, Pensacola, FL, June 1988.

17. Neuman, T. Cross-validation of the defense mechanism test by the criterion of passing or failure in basic military flight training. Report No. FTD-HC-23-1150-72, Wright-Patterson AFB, OH, August, 1972.

18. Jackson, D.N. Personality Research Form Manual, Research Psychologists Press, Goshen, NY, 1974.

19. Anastasi, A. Evolving Concepts of Test Validation. Annual Review of Psychology, 37, 1-15, 1986.

20. Skinner, H.A., Jackson, D.N., and Rampton, G.M. The personality research in a Canadian context: Does language make a difference? Canadian Journal of Behavioral Sciences, 8, 156-168, 1976.

21. Joaquin, J.B. The personality research form and its utility in predicting undergraduate pilot training performance in the Canadian Forces (Working Paper 80-12). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit, 1980.

22. Valentine, L.D. Navigator-observer selection research. AFHRL-TR-77-36, Air Force Human Resources Laboratory, San Antonio, TX, May 1977.

23. Retzlaff, P.D. and Gibertini, M. Air Force pilot personality: hard data on the right stuff. Multivariate Behavioral Research, 22, 383-399, 1987.

24. Cattell, R.B., Eber, H.W., and Tatsuoka, M.M. Handbook for the Sixteen Personality Factor Questionnaire. Institute for Personality and Ability Testing, Champaign, IL, 1970.

25. Bartram, D. Development and evaluation of Micropat Version 4. Ministry of Defense, London, and Ergonomics Research Group, University of Hull, TR-184, 1986.

26. Cattell, R.B. and Stice, G.F. Handbook for the Sixteen Personality Factor Questionnaire. Institute for Personality and Ability Testing, Champaign, IL, 1962.

27. Fleischman, H.L., Ambler, R.K., Peterson, F.E., and Lane, N.E. The relationship of five personality scales to success in naval aviation training. NAMI-968, Naval Aerospace Medical Institute, Pensacola, FL, 1966.

28. Lester, L.F. and Bombaci, D.H. The relationship between personality and irrational judgment in civil pilots. Human Factors, 26, 565-572, 1984.

29. Rotter, J.B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs: General and Applied, 80, 1-28, 1966.

30. Wichman, H. and Ball, J. Locus of control, self-serving biases, and attitudes towards safety in general aviation pilots. Aviation, Space, and Environmental Medicine, 54, 507-510, 1983.

31. Alkov, R.A., Gaynor, J.A., and Borowsky, M.S. Pilot error as a symptom of inadequate stress coping. Aviation, Space, and Environmental Medicine, 56, 244-247, 1985.

32. McGuire, F.L. Personality factors in highway accidents. Human Factors, 18, 433-436, 1976.

33. Jenson, R.S. and Benel, R.A. Judgment evaluation and instruction in civil pilot training. TR-RD-78-24, Federal Aviation Administration, Springfield, VA, 1977.

# ACTIVATION – POSITIVE AND NEGATIVE EFFECTS OF THE ALARM SYSTEM IN THE BRAIN

by

Prof. Holger Ursin
Dept of Physiological Psychology
University of Bergen
5000 Bergen, Norway

## Activation and aviation

There is hardly any human activity where the level of wakefulness may have so dramatic effects on performance with consequences for life and death, as aviation and other forms of rapid transportation. All aspects of the problem area seem to be present: boredom and lack of attention during long periods of routine operation, bursts of activity at top performance level, and the possibility that there are long term effects which may be harmful in the long run. Numerically land transport is by far the most dangerous and costly operation, counted in lives lost or invalids produced. However, the concern in this paper is aviation, but the factors involved are general psychological and physiological principles valid for many types of activity. The mechanisms are actually biologically general as well, and some of the relevant data derive from animal experimentation.

## Specific and nonspecific responses

There is a reasonable degree of agreement among researchers that whenever there is a change in the environment, two types of responses are seen in any organism. The first is the specific ("unconditioned") responses that are specific to that particular change. Pupils contract to light, urine is concentrated when there is a shortage of water. These mechanisms are prewired, and stereotyped. They are found in primitive organisms, and do not require much of the brain's capacity to treat information. The mechanisms are necessary for survival, but do not really give the individual many degrees of freedom with regard to the environment.

The second set of responses are complex, and represent behavioral responses which contribute to the adaptation to change in the environment. These mechanisms are often typical for one individual, and represent individual adaptations to the environment. They represent learning, or individual storage of information about relationships between responses and stimuli, or between stimuli. The responses vary from a general alarm system to specific, almost automatic responses; from apparent senseless moving around in desperate situations, to moving in the chair when the pressure becomes too much for sensitive parts of the body. The responses represent a continuum of behavioral responses emitted under very high activation of endocrine, autonomic and immune responses, to perhaps the same responses emitted under no or only very limited change of the endocrine, autonomic, and immune systems. There may not be that many different actions to be observed in the performance of a pilot landing under circumstances where he or she is uncertain as to the status of the landing gear, but the subjective experience and the physiological state are very different from a routine operation. There are also situations where no behavior is emitted, but when there are strong responses in the endocrine, autonomic, and immune systems. The reasons for concern with the high arousal or activation is twofold. Such responses require more information capacity of the brain. Second, it may be reported as "stressful" or unpleasant, and there is a possibility that it may be related to disease risk.

## Activation and homeostasis

There a reasonable degree of agreement that when there is a discrepancy between what the organism is set for regarding an important variable, and what really exists (actual value of the same variable), there are also two types of responses. The first is the general alarm response described above. This is in principle a non-specific response affecting all somatic systems, directly or indirectly. The second type or class of responses are specific to the situation and the experience of this particular individual.

There is no consensus to what the general alarm system should be called, and some may even question its existence. Even so, the general textbook view is that we accept that brains are equipped with a central, brain stem and diencephalic system that regulates the level of wakefulness. There is also an interacting system that regulates the different types of sleep. The general activation system is dependent on the reticular formation, and on specific ascending systems, also located in the brain stem. The relationship between this system and what people refer to as the stress response is more controversial. This author is unable to see any differences in the responses described as stress responses, and those that are concomitant with the general arousal or activation seen when a subject goes from one state of wakefulness to a level of higher wakefulness (Ursin 1978, 1988).

There is universal disagreement to the existence of specific responses, beyond what each individual may learn. Some claim specific pathways in the brain from one type or class of stimuli and specific parts of the alarm system, some claim specificity between particular behavioral strategies or response systems and particular hormones. This problem area is closely related to personality traits, and, therefore, relevant for selection, and will be treated separately below.

## Activation

The strong correlations between the subjective feeling of wakefulness versus drowsiness and the corresponding changes in the electroencephalogram (EEG) (Jones 1981), were the basis for the original classification of the EEG by Berger (1930). This was also the basis for the neurophysiological activation concept as we know it today. The essential aspect for the use in the present context is the relationship between this central state and vegetative, endocrine and immunological processes. The changes produced in these subsystems are to be regarded as parts of the total response. The feedback from these responses to the brain is mainly a positive feedback system, producing more activation. There may also be specific effects on consolidation processes (deWied 1974) and later behavior. The various subsystems are contributing parts, but none of them are essential elements. Changing one or several of these systems may dampen the response, but does not eliminate the response.

All endocrine systems, the autonomic systems, and the immune system, have been shown to be "sensitive" to environmental changes. This has also been shown for the biochemical composition of the brain itself. This is a direct consequence of these parts of the body being essential elements in the general ability to adapt to changes in the environment, which is a condition for survival. It should not be surprising that this is due to brain control over these mechanisms, or that psychological factors are of decisive importance.

## Activation - the stress response

Even if there seems to be no principle difference between the general activation response, and what happens when people report stress, it is important to note that activation also occur during states that are not reported as stress. All stress responses are activation responses, but not all activation is experienced as "stress". Theoretically, stress reports, and stress responses, seem to occur when the brain registers a lack of something it regards as necessary or at least desired. The stress response, therefore, is the general alarm system in the brain and in the body for this state. It elicits somatic and behavioral responses which tend to eliminate this imbalance, just as any other homeostatic mechanism. Unfortunately, stress is also used for the stimuli eliciting this state, and for the subjective experience of this state, and that disease is assumed to result from this homeostatic mechanism. The problems and ambiguities of the term will have to be dealt with. In particular, we must discuss why this perfectly normal and adaptive response have obtained such a bad reputation for producing disease, and for interfering with normal and adequate behavior in difficult situations.

Within control theory, stress responses, and all types of activation, occur whenever there is a discrepancy between a set value and the real value (actual value) for that particular variable (Ursin 1978). This is only a reformulation of the basic law of physiology, and physiological psychology, which is that the body regulates itself according to homeostasis. In spite of changes in the external environment, we keep our internal environment constant (C. Bernard- 1813-78). The concepts of selfregulating systems in control theory and information theory came much later, about 1940-1950. Physiological "theory" with its homeostatic principles antedates this by 70-80 years. The recognition of a general activation system in the brain was first attached to emotions and sympathetic discharge. Then, in 1949, the existence of a general activation system (Moruzzi and Magoun 1949) changed our conceptualizations of central nervous function, and, also, of psychology. This alarm system drives the brain and the organism to action, until the solution is found, or a new set value has been established, or the brain switches to a different motivational system (Ursin 1988).

## Psychological deactivating principles

There are many terms for the various psychological mechanisms that interfere with these response systems. We believe that there is considerable amount of consensus for the principles, but a lack of agreement and consistency in the nomenclature used. Both in animals and in man there is a high level of consistency in the findings that uncertainty, lack of information, and lack of control over the environment produce alarm states. Conversely, presence of control and information, clear and salient safety signals, and high levels of performance reduces the internal state even if the same behavior is emitted. Performance may actually be very efficient even under very hazardous conditions with very low levels of endocrine or autonomic activation, if the individual has acquired a positive expectancy to the outcome of this situation. The disagreement, or inconsistency, in the literature, is due to differences in theoretical background and nomenclature. The empirical base is the same.

In the following the relevant terms will be defined in terms of expectancy, in particular response outcome expectancies. The underlying theoretical assumption is simply that brains are able to store information, and that the stored information has the nature of "expectancies". It seems to be an economical and parsimonious position that brains store relationships either between stimuli, or between responses and stimuli. The first phenomenon may be referred to as stimulus expectancy, the second as response outcome expectancy (Bolles, 1972).

When brain activity is to be regarded in a more total biological context than in traditional learning theory, it is necessary to use terms like attention, preparedness, or simply expectancy. To perform complex acts like catching a prey, the predator must direct its movements to where the prey is expected to be in the next time interval. When the brain has established that something proceeds something the brain may be said to expect the second stimulus once the first stimulus has been presented. This is a reformulation of what happens during classical conditioning. Brains may, of course, also register what happens after a certain response has been emitted. This is the essence of instrumental conditioning, which therefore may be referred to as response outcome expectancy.

Defense.

Any stimulus that is expected to be followed by an aversive event (classical conditioning) has the potential of becoming a stressful or stress-producing stimulus. There are other conditioned factors that may interfere with the production of a stress response, which mainly have to do with response expectancies (see below). But in animals the signals of unavoidable, uncontrollable aversive events produce stress responses with a high level of predictability. Similar findings have been made for humans, both in the laboratory, and in real life situations.

Humans seem to have one more mechanism available. They may initially perceive a threat signal, but then distort and deny the true or probable consequences of that signal. In its most primitive forms perceptual defense may block threat signals from producing stress responses, but this may also have as a consequence inadequate behavior in dangerous situations (Kragh 1960, Værnes 1982). Strong tendencies to use such strategies seem to have the characteristics of a personality trait, a stable characteristic of that person, and is being used as a negative selection criterion for dangerous occupations in Scandinavia (Kragh, 1960, Værnes et al this volume). Lazarus (1960) comprised defense in his coping concept, while other authors make a firm distinction between coping (healthy) and defense (non-healthy), for instance Haan (1978). We believe that this distinction is important.

This ability may be very efficient in producing safety and reduce the physiological responses to threat, but the price may be rather inadequate behavior. The concept is an important part of psychodynamic thinking. It may also be used in a more restricted or conservative sense, relating to the perceptual distortions of reality or true stimulus contingencies, as is the case in the present paper. In the material this review is based on defense has been operationalized either as distortions and misinterpretations in a tachistoscopic, perceptual task (Kragh 1960, Værnes 1982), or as scores in a paper-and-pencil test which is constructed based on clinical material (Plutchik, Kellerman and Conte 1979). Værnes and collaborators review recent experience with these tests elsewhere in this volume, and recent reviews of the test from Britain (Kline 1986) are available. The Kragh test taps a form of misinterpretation and misdirection of attention and perception, which operates when a subject is perceiving real danger. There is no established relation to performance during ordinary training sessions (Ursin, Baade and Levine, 1978, Værnes 1982).

Spatial and analytical functions in perception are mediated by separate brain substrates. Interference between these cognitive modes degrades pilot performance (Sterman et al 1988). It is not known whether these two perceptual phenomena that interfere with pilot performance are related. The personality trait with high defense may also be related to the personality characteristics discussed by Foushee in this volume. These possible relations have not been tested yet.

Coping

The term coping is, as mentioned earlier, used in two different meanings. We will refer to coping strategies for the behaviors that are being used in order to face a challenge or threat. These strategies may be executed both in high and low levels of physiological activation states (see below). Coping in this paper will refer to established positive response outcome expectancies. Coping, therefore, is the result of a learning process. An individual expects a positive outcome with a high subjective probability (perceived probability), with a high level of learning (habit values), and the results of these acts are regarded by the individual as highly attractive (high positive affective value). Coping, therefore, means a positive response outcome expectancy. When given this definition, coping predicts a reduced activation level in the physiological state of the individual. This has been demonstrated both in animals ( Coover, Ursin and Levin 1973), and in humans ( Ursin, Baade and Levine 1978). Absence of coping (see below), accordingly, may result in continuous activation and secondary pathological states (Ursin and Murison 1983).

## Psychological activating principles

It should follow from the above that the absence of coping, or control and predictability, should produce activation. The extreme lack of information may be used as a definition of anxiety (see Ursin 1978 for discussion), and information reduces anxiety and uncertainty. Both animals and humans prefer information to uncertainty, predictability to unpredictability, and show decreased activation when this is obtained (Goldman, Coover and Levine 1973). There are particularly two states that have received much experimental attention in the last decade, that both may be redefined as special cases of response outcome expectancies.

## Helplessness

Helplessness refers to psychological situations where the individual does not see any relationship between the responses available to the individual, and the possible outcome of the situation. This definition is, in our opinion, reasonably close to the original offered by Overmier and Seligman (1967). Within our theoretical framework, helplessness exists when the subject has learned (with high habit values) that there seems to be a very low probability that responses will lead to any result at all. This state has been established mainly in experimental situations involving stimuli with high negative value, but it may exist also in situations with positive affective values. The effect on the internal state under such circumstances have not been measured. The state under the threat of highly negative outcomes unless rectified by the individual is high arousal.

## Hopelessness

Hopelessness in this chapter means that the individual has acquired a high perceived probability that available responses will lead to negative events, that is stimuli with a high negative affective value, or, simply, highly aversive. Whatever the subject does, something, or most things, go wrong. This state is more related to the guilt-ridden depression often typical for human depressive states. Helplessness has often been used as a theoretical model for depression (Seligman 1975), however, in our opinion, hopelessness may be a better model (Pociuk, Breen and Lussier 1976).

## Activation theory

Coping, helplessness and hopelessness are regarded as states elicited by a certain stimulus situation, i.e. a stimulus expectancy. These states are not only cognitive states, they also have well defined consequences for the physiological state of the organism. We have chosen to explain this within the framework of activation theory, as it is known in neurophysiology (Moruzzi and Magoun 1949, Lindsley 1951). A discussion following a paper by Vanderwolf and Robinson (1981) illustrates what is still the current status of the concept. In the present context, we assume simply that activation is a process in the central nervous system (CNS) which increases the activity in the brain from one level to a higher level, and which may maintain this level. This happens when we wake up, and is basically the same response we use when responding to novel stimuli or when facing a threatening challenge or an interesting problem. The emphasis of this concept is on changes in these levels, and maintenance of them.

The presence of activation may be regarded as the driving force behind solution of problems. Activation may be regarded not only as an alarm system, but the driving force that makes an animal or a human act to reduce needs. Activation, therefore, is in this context an essential element in the total adaptive system of the organism (Ursin 1988). It is not to be accepted as an atavistic mechanism no longer suitable for civilized man (Levi 1972, 1987, Charvat, Dell and Folkow 1964). Feelings of anxiety, unrest, stress and conflict are not necessarily evils to be dampened by psychopharmacological interventions, they may be an adequate response to stimuli requiring full attention and integrated action for solution, and subsequent reduction of the activation.

## Performance consequences

The curvilinear relationship between arousal (activation) and performance (Malmo 1966, Duffy 1972) is well known to all psychology students, both from textbooks, and sometimes from their own experience. It is, however, not that easy to reproduce, and occurs only under certain circumstances (Hockey 1984). The classical position is that performance is optimal for one value of activation, and becomes gradually impaired on each side of this point, gradually decreasing with the distance from the optimal value. Activation has also been suggested to be homeostatically controlled itself, at least in the awake subject.

The curvilinear relationship is important for preparing crews for optimal performance in critical situations. Strong levels of emotions, for instance aggression or fear, and high defense mechanisms must be assumed to interfere with communication and performance in critical situations. The selection criteria discussed by Foushee and by Værnes and collaborators in this volume address such issues. High levels of coping, with the accompanying trust in ones own performance, must be supposed to be a positive contribution to the same variables.

The reason high affect (and low coping) may interfere negatively with performance in critical situation is best explained by the limited channel capacity theory of central nervous function (Broadbent 1971, Hamilton 1975).

## Psychosomatic consequences

When coping has been established, there is still a shortlasting activation which has been referred to as "phasic" activation, in contrast to the general, and longer lasting activation seen in the non-coping subject (Ursin 1978). The phasic activation in the coping individual is characterized by epinephrine release (not norepinephrine) (Hansen et al 1978), pulse rate increases (Strømme et al 1978), and a modest but significant rise in plasma levels of testosterone (Davidson et al, 1978). Arnetz (1984) produced this type of activation in elderly as a response to an intervention program, and pointed out the desirable anabolic aspects of this type of activation.

The tonic activation seen in the non-coping individuals, or during the first period when a subject is faced with a taxing or threatening situation, is not directly related to pathology, either. In a weak individual this may be too much, the drop to make the flow-over effect, but still should not be accepted as the main cause of the pathology. This aspect of "stress" may be an important rational for the relief part of nursing. Medicine has had this concept as an integral part of all treatment of the sick since Hippocrates (460-370 BC). The same "relief" idea may also be one of the underlying concepts that lead Selye to postulate his general stress theory, where loads or "stressors" are assumed to have an almost additive effect in a very non-specific fashion to produce the general appearance of a "sick" person (Selye 1950).

However, as a theory for the occurrence of psychosomatic disease this is still not anywhere near an acceptable pathophysiological model. How can these normal, adaptive, and physiological mechanisms produce or contribute to pathogenesis in an otherwise healthy organism?

In a previous paper, Ursin (1978) has suggested that pathophysiology may occur as a consequence of sustained, tonic activation in genetically or otherwise predisposed individuals. He found no evidence for, nor any reason to postulate, any specific links between any special stimuli, conflicts or situations and specific types of pathology, as suggested by the traditional psychodynamic writers (e.g. Alexander 1950).

There is no single, common mechanism. Sustained high norepinephrine levels might produce high blood pressure in individuals predisposed for that disease. Sustained high levels of plasma cortisol may also contribute to elevation of blood pressure, but may also be related to changes in immune functions. Elevated free fatty acids, which is another concomitant of high tonic activation, may be related to cardiovascular pathology. For many of the consequences or parts of the sustained activation state the prolonged status out of the normal equilibrium may be postulated to contribute to pathophysiology, but in interaction with other pathophysiological factors. The psychosomatic factors are risk factors, as are most or all other factors producing disease. In other words, this is a psychosomatic theory which is well within and compatible with the ordinary, multicausal ideas of the origins of disease.

References

Alexander, F. (1950) Psychosomatic Medicine. New York, W.W. Norton.

Arnetz, B.B. Thesis, University of Stockholm, 1984

Berger, H. (1930) Uber das Elektroenkephalogramm des menschen. Journal of Physiology and Neurology, 40: 160-179.

Bolles, R.C. (1972) Reinforcement, expectancy and learning. Psychological Review, 79: 394-409.

Broadbent, D.E. (1963) Differences and interactions between stresses. Quarterly Journal of Experimental Psychology, 15: 205-211.

Charvat, J., Dell, P., and Folkow, B. (1964): Mental factors and cardiovascular disorders. Cardiologia, 44: 124-141.

Coover, G.D., Ursin, H. and Levine, S. (1973): Plasma corticosterone levels during active-avoidance learning in rats. Journal of Comparative and Physiological Psychology, 82: 170-174.

Davidson, J., Smith, E.R., and Levine, S. (1978): Testosterone. In Ursin, H., Baade, E., and Levine, S. (eds.) Psychobiology of Stress. A Study of Coping Men, New York, Academic Press, 57-62.

deWied, D. (1974) Pituitary-adrenal system hormones and behavior. In F.O. Schmitt and F.G. Worden (Eds.): The Neurosciences. Third Study Program. Cambridge, The MIT Press, 653-666.

Duffy, E. (1972) Activation. In N. Greenfield and R. Sternback (Eds.): Handbook of Psychophysiology. New York, Holt, Rinehart & Winston, 577-622.

Folkman, S. (1984) Personal control and coping processes: A theoretical analysis. Journal of Personality and Social Psychology, 46: 839-852.

Goldman, L., Coover, G.D., and Levine, S. (1973): Bidirectional effects of reinforcement shifts on pituitary-adrenal activity. Physiology and Behavior, 10: 209-214.

Hamilton, V. (1975) Socialization anxiety and information processing: A capacity model of anxiety-induced performance deficits. In I.G. Sarason and C.D. Spielberger (Eds.): Stress and Anxiety. Vol.2. Washington DC, Hemisphere (Wiley), 45-68.

Hansen, J.R., Støa, K.F., Blix, A.S., and Ursin, H. (1978): Urinary levels of epinephrine and norepinephrine in parachutist trainees. In: Ursin, H., Baade, E., and Levine, S. (eds.): Psychobiology of Stress. A Study of Coping Men. New York, Academic Press, 63-74.

Hockey, G.R.J. (1984) Varieties of attentional states. The effects of environment. In R. Parasuraman and D.R. Davies (Eds.): Varieties of Attention New York, Academic Press.

Jones, B.E. (1981) Understanding the physiological correlates of behavioral state as a constellation of events.The Behavioral and Brain Sciences, 4: 482-483.

Kline, P. (1986) Personality inventories. In S.E. Newstead, S.H. Irvine and P.L. Dann (Eds): Human Assessment: Cognition and Motivation. Dordrect, M. Nijhoff, 109-121.

Kragh, U. (1960) The defense mechanism test: A new method for diagnosis and personnel selection. Journal of Applied Psychology, 44: 303-309.

Lazarus, R.S. (1983) The costs and benefits of denial. In S. Breznitz (Ed.): The Denial of Stress. New York, International Universities Press.

Levi, L. (1972): Stress and distress in response to psychosocial stimuli. Stockholm, Almquist & Wiksell.

Levi, L. (1987) Psychosomatic disease as a consequence of occupational stress. In: Kalimo, R., El-Batawi, M.A., and Cooper, C.L. (eds.): Psychosocial factors at work and their relation to health. Geneva, World Health Organization, 78-91

Lindsley, D.B. (1951) Emotion. In S. Stevens (Ed.) Handbook of Experimental Psychology. New York, Wiley.

Malmo, R. B. (1966) Some clinical origins of the activation concept. In C.D. Spielberger (Ed.): Anxiety and Behavior. New York, Academic Press, 157-177.

Moruzzi, G. and Magoun, H.W. (1949) Brain stem reticular formation and activation of the EEG. Electroencephalography and Clinical Neurophysiology, 1: 455-473.

Overmier, J.B. and Seligman, M.E.P. (1967) Effects of inescapable shock upon subsequent escape and avoidance responding. Journal of Comparative and Physiological psychology, 63: 28-33.

Prociuk, T.J., Breen, L.J., and Lussier, R.J. (1976) Hopelessness, internal-external locus of control and depression. Journal of Clinical Psychology, 32: 299-300.

Seligman, M.E.P. (1975) Helplessness: On Depression, Development and Death. San Francisco, Freeman.

Selye, H. (1950) Stress. Montreal, Acta Incorporated.

Sterman, M.B., Dushenko, T. Schummer, G., and Smith, J.C. (1988) EEG correlates of flight performance. In Electrical and Magnetic Studies of Brain Function in Relation to Aerospace Medicine, AGARD, 1988, Trondheim.

Strømme, S.B., Wikeby, P.C., Blix, A.S., and Ursin, H. (1978): Additional Heart Rate. In Ursin, H., Baade, E., and Levine, S. (eds.): Psychobiology of Stress. A Study of Coping Men. New York, Academic Press, 83-89.

Ursin, H. (1978) Activation, coping, and psychosomatics. In H. Ursin, E. Baade and S. Levine (Eds.): Psychobiology of Stress: A Study of Coping Men. New York, Academic Press. 201-228.

Ursin, H. (1988) Expectancy and activation: An attempt to systematize stress theory. In D. Hellhammer, I. Florin and H. Weiner (Eds.): Neurobiological Approaches to Human Disease. Toronto, Hans Huber.

Ursin, H., Baade, E. and Levine, S. (Eds.) (1978) Psychobiology of Stress: A Study of Coping Men. New York, Academic Press.

Ursin, H. and Murison, R. (Eds.) (1983) Biological and Psychological Basis of Psychosomatic Disease. Oxford, Pergamon Press.

Vanderwolf, C.H. and Robinson, T.E. (1981) Reticulo-cortical activity and behavior: A critique of the arousal theory and a new synthesis. The Behavioral and Brain Sciences, 4: 459-514.

## DISCUSSION

SIOMOPOULOS: We have listened to a number of papers in our effort to understand human behaviour. I think we must keep in mind that there are several levels of discourse which should be kept separate. One level is called psychosocial or the level of the person. The other level is that of information processing which Dr Damos was talking about, and she kept nicely on the level and she gave an excellent presentation. Another level is that of the organism, neurophysiological, humoral or whatever. Even at that level there are some variables that may be indices of behaviour, even the EEG. Then there is another level that psychoanalytic thinking has claimed, the intra-psychic level.

Now, these levels have been mixed up in such a way that there is some confusion. With regard to defence mechanisms, a concept that is purely psychoanalytical, we talk about high defences and low defences. Yet since the time of Anna Freud, who systematised whatever Freud has said about it, even the concept of defence mechanisms has been completely revised in psychoanalytic thinking. The concept of psychic energy has been rejected from psychoanalytic theory. It is not really advisable to utilise this kind of terminology in such a haphazard way because, if we do, then we do not really understand what we are talking about. I would like to voice this warning because I really got very confused about some of the things presented. I think we have to stick very clearly to the levels we study and be clear about whatever methodological approaches we are using.

URSIN: I would start with agreeing with your premises but I strongly disagree with the conclusions. If our task is to understand what goes on in the human individual, there is no way I can accept it when you come to me with this level and that level. I am dealing with live human beings, and they do not come with levels. They come as live human beings. They have psychological backgrounds with psychological traits, changed immune levels and organic disease. Or are they just concerned about getting an organic disease? So it is my task then to go through all levels and see how I can use this information. If we had been forced to stick to traditional schools we would never have seen that whatever is measured in the "Crack" test has very strong and consistent correlations with cortisol levels. Of course, it goes across everything. We even did a factor analysis on the endocrine responses. No sane endocrinologist would ever do that. Actually most sane endocrinolgists do not know what factor analysis is. There is some reason to look into that. There are orthogonal factors in the endocrine response and these factors correlate with personality.

You ask for definitions and that is why I said that when I am talking about defence I am talking about stimulus expectancy, so I am using learning theory terms for whatever is measured by the Crack Test. The Crack Test measures some kind of personality trait that effects stimulus expectancies and I call that defence, I do not care one bit about the psychoanalytic theory that was behind the concept in the first place. Actually I will go one step further I will even say that this trait, that comes from psychodynamic thinking, has an organic basis which is totally against anything in psychoanalytic thinking. I will go even one step further. I think that the data Barry Stirling is getting these days, although not represented here, shows that there is some kind of lack of co-ordination between the right hemisphere and the left hemisphere in people who are not able to transfer from analogue data to digital data and back again. I have a very strong suspicion that this type of thinking, this type of phenomenon, may be related to perceptual phenomena which are deeply ingrained and which do have something to do with development of personality traits. The genetic background for this may be very complicated, but that part of it we do not know. The strongest data we have is from the Crack Test where a percept is built up within fractions of a second. That this relies on a neurophysiological mechanism is obvious, but it could also be related to personality traits. This is an interesting possibility which I think we should look into it.

SIOMOPOULOS: I was misunderstood and I want to clarify something. I did not deny any of the rights of Professor Ursin to believe whatever he wants to believe and to discard whatever he wants to discard. I just wanted to voice a small warning. When we try to talk from level to level let us try to make sure that the bridge we build from one level of discourse to another level stands on a sound, scientific and methodological foundation. That is all, nothing more than that. Everybody has a right to believe and discard whatever he wants. Building bridges between levels of discourse is not an easy task. We have to make sure that at least we know exactly how to connect the levels.

PSIMENOS: I was fascinated by the way you handled the whole phenomenon of a human being lit in its extreme. Pilots are similar to some other groups. I am dealing with another group of people who live beyond death, namely, those on dialysis. In them you can study the human being in its extreme course towards the unknown; they are going out and meeting with death. They are surpassing their programmed abilities and creating, perhaps, new reflexes, so it is not surprising that so many views about what to look for and how to select people exist. There will probably be a new generation of human beings that are so daring and so able to perform in an alien environment that is totally unknown so far. It is the speed of events that changes things. Homeostasis, such as you mentioned, is an excellent example of keeping stable in a changing environment. If things happen very quickly perhaps homeostasis, which is synonymous with life, has no time to take place. The whole complex of changes, many more than the ones you described, should really be studied.

URSIN: I think I should have been arrested for my use of the term homeostasis because I made an issue of talking about homeostasis as the regulating principle, but I have deviated from that twice. Once in the activation response itself which is a non-homeostatic concept. It is an out-of-bounds response which is built into the system. In order to modify homeostasis theory you have to say that there is an alarm system that operates when these discrepencies exist. One is willfully, as it were, taking the system out of homeostatic balance and creating a situation which will drive the individual to the correct behavioural responses in order to re-establish the situation. So now you have to allow for a homeostatic gate, as it were, that will say this is minatory after half an hour type of event that happens. Also when you kick off the activation it is a very strange response. It has a time course which makes it very difficult to handle in a unitary manner, because activation 5 s after the start is different from that occurring at 10 min.

Now I also postulate that there is a price for sustained activation, another non-homeostatic principle. If you are going to accept a psychosomatic theory then you will have to accept that there is the possibility of getting the homeostatic system out of balance. On the other hand, if you were a molecular biologist you would say that it is probably happening because when you stress the system you are changing the receptors, and if you have a dynamic picture of receptors then you will have to rethink it. But generally speaking we are back to a homeostatic principle with some rather important modifying factors. It is, however, a rather complex issue.

BILLINGS: I think you are being unfair to yourself. Yes, the alarm response represents a departure from homeostasis in the short term, but continued alarm responses over a considerable period of time are necessary for the establishment of a stable homeostatic mechanism in an adult. We have seen that in rats and in humans over many, many years. It takes alarm responses to learn how to respond to alarms.

URSIN: It may take alarm responses to be able to teach individuals what to do during an alarm. That is back to our discussion on Tuesday on the state-dependency part of learning.

SCHILDER: With cancer patients sometimes we notice that after diagnosis of a life threatening event there is a quite remarkable and sudden change from say helplessness and defence avoidance into aggressiveness and prognophobic behaviour. In such situations one notices definite changes in the personality of the person. The person himself does not notice it. It is a sort of sudden change in set values that you might explain by Gatastrov's or Gale's theory.

URSIN: I do not know how to handle such a shift in strategy, as it were, and I am not sure that thinking about set values really contributes. I think it may be more a shift in state. A change in the motivational situation brings about a shift in the total repertoire you are trying to attack with. I do not think the answer is hidden in the general control figure, it is hidden in the fact that the situation is changed, therefore you shift strategy.

SCHILDER: Do you observe within the strategy surge, if I may use the word, the sudden shift from an inability to behave in a very useful way to a way of, let's say, an almost higher organisational reaction, especially in the face of life threatening situations which pilots or patients can both encounter?

URSIN: There are two aspects of that, if I understand you correctly. One is the shift in strategy when you have a shift in the perceptual situation. This is just one result of all the things that happen when you go from a high instrumental face and you suddenly realise that something is wrong and that your ordinary skills are no longer adequate. Then you go from one state to a different one. That was what Billings and I were talking about. In order to be able to handle that you have to realise that a pilot is in a different type of state, a totally different type of state. I think that may also be what is going on in your patient. The other thing is that when you attack a new type of problem, and you realise that you have to shift strategies, this may not be a danger signal but rather a positive signal given so that you realise your worse fears have come true. Now you have to handle it differently, so you start with aggression. That may not be such a good idea, but you are at least starting new programs. One of the things one attempts in psychotherapy, for a person who has tried the same strategies over and over again, is to inculcate some other strategies. You have to offer some kind of alternative strategy in order to try to change the situation. Get the helpless person out of helplessness, and if aggression is the price then I will take aggression. But get the person out of this "nothing works, nothing helps" state.

RODRIGUEZ: There is only one question about the method. I have seen that the cortisol and catecholamine levels were increased by 250 to 300% over the base levels. Speaking about catecholamines; did you determine epinephrine or norepinephrine levels in plasma or urine? We have experienced problems working with such hormones in other situations. Working with plasma catecholamine is difficult just because it is secreted in peaks. I must stress that the finding of an increase from normal by 300% is a really big change in response to the stimulus.

URSIN: I think I simplified the slide that I showed you. The starting level was 100%, so you are right, the level went up to 200%. I do not think I had any epinephrine values on that slide, I had free fatty acids. When we measured epinephrine and norepinephrine we did that after really miserable attempts on the plasma so we went back to urine. For this type of research I think I will go as far as to recommend very strongly that one stays off the plasma. The plasma is awful. The plasma is just a gruesome, inefficient way of measuring muscle activity. If you pump the arm you empty all the epinephrine from the muscles. After all the attempts to be really sophisticated it was useless, so we went back to determining urine levels.

# THE TRIALS AND TRIBULATIONS OF RAF DEFENCE MECHANISM TESTING

DR G J WALKER-SMITH
SCIENCE 3(AIR)
MINISTRY OF DEFENCE
LACON HOUSE
THEOBALDS ROAD
LONDON WC1X 8RY
UK

The Defence Mechanism Test (DMT) is a projective personality test that was devised in Sweden, in the 1950s. It was designed to identify how individuals cope with a threat and to assess what defence mechanisms are used to protect the individual in a stressful situation. The test was developed for use in the Royal Swedish Air Force (RSwAF). Swedish validity studies have shown that the DMT can predict training wastage and pilot error flying accidents. Consequently the DMT has been used since 1970 for RSwAF pilot selection. The test has also been used for Air Force pilot selection in Norway, Denmark, Greece and the Netherlands and is undergoing trials in other countries. Since 1976 empirical investigations have been carried out to test the validity of using the DMT to select Royal Air Force (RAF) pilots. Owing to methodological inadequacies, early RAF trials proved inconclusive. However, in 1984 a DMT trial was set up where the Swedish method of testing was followed. DMT scores were collected from a sample of 253 pilot trainees and their flying training results and flying accident involvement are being monitored. So far this DMT trial has shown that the DMT scores fail to predict flying training performance. The discrepancy between the RSwAF results and the RAF findings is considered.

The Defence Mechanism Test (DMT) is a projective personality test that was devised in Sweden, in the 1950s, by Ulf Kragh. The test was originally constructed for the diagnosis of defence mechanisms and for personnel selection. It was developed principally for the selection of pilots, particularly military pilots, although its applications are now widespread.

Essentially the DMT involves the repeated subliminal presentation of Thematic Apperception Test (TAT) type pictures. The stimulus pictures, which depict a central hero figure and a threatening peripheral person, are displayed via a tachistoscope (a devise for presenting pictures for very short durations), and the exposure durations are increased at each presentation. After each exposure, the subject is required to sketch and describe what he has seen. The analyses of the results involves examining the progression of misperceptions, or deviations from the stimulus picture. These deviations are then interpreted in psychoanalytic defence mechanism terms.

The percept-genetic model of perception and personality, the psychoanalytic theory of defence mechanisms, and the theory and practice of projective techniques were used by Kragh as the rationale behind the design of the DMT. According to percept-genetic theory, perception is not merely a reflection of the outside world. Instead percepts develop through the interaction between the perceiver and the stimulus. Thus an individual's personality influences the way in which the outside world is perceived.

Phenomenologically, percepts appear to come about as soon as a stimulus is confronted but percepts do not appear instantaneously - they are the end product of a process that extends over a brief time period. The end product of the perceptual process in percept-genesis is referred to as the concluding, or consensus, C-phase. The earlier stages that the perceiver is not usually aware of are termed the preparatory, P-phases.

Tachistoscopic techniques are used to reveal the perceptual processes that lead to the appearance of the C-phase. By presenting a stimulus briefly, just below the threshold of the subject recognizing that anything is being presented, and then successively increasing the exposure duration until the consensus meaning is reported, the P-phases are prolonged and the final C-phase is prevented from appearing instantaneously. The succession of reports is treated as a reflection of the perceptual process. Furthermore, the perceptual processes are believed to represent the perceiver's life-history in a parallel fashion. The percept-genesis corresponds sequentially to significant events in the individual's life-history - early experiences are revealed in early P-phases and later experiences are shown in subsequent ones.

The DMT was designed to reveal perceptual defence and adaptation processes. The subject is assumed to identify himself with the hero. The threat in the picture initiates a subliminal anxiety reaction and the subject either reports the threat or the stimulus evokes defence mechanisms. According to psychoanalytic theory these mechanisms are stable phenomena and they transform the threat to make it non-threatening. The DMT elicits the classical psychoanalytic defence mechanisms, such as repression, isolation, denial, reaction formation, identification with the aggressor, intro-aggression, introjection, projection and regression. The distortions in the DMT responses indicate which defence mechanisms are operating. For example a line drawn between the hero and threat figure may indicate "isolation" or a happy threat figure might signify "reaction formation".

Where defence mechanisms are constantly at work people have difficulties in resisting stress. In stressful situations the use of defence mechanisms limits the amount of energy available for reality-testing and adaptation. In military flying, where the job itself induces anxiety, mistakes will occur if the pilot makes faulty or dangerous decisions because he cannot correctly identify threats and deal with them effectively. Thus it is assumed that in aviation training people with strong defence mechanisms would be accident prone, owing to deficiencies in reality-testing, and would eventually fail in training or have accidents whilst flying operationally.

The involvement of the Royal Air Force with the DMT has extended over many years. In the 1960s RAF aviation psychologists heard about Kraghs work. Their interest was stimulated further, in 1971, when Thomas Neuman visited Cambridge and presented a paper that described how the DMT predicted military pilots' flying performance (1). The results were impressive. With the introduction of the DMT as part of the Royal Swedish Air Force selection procedure, in 1970, the pass fail ratio among students who entered basic flying training was reversed from 40/60 to 60/40. Consequently for the same training output, few students needed to be trained.

The first exploratory RAF DMT study was started in October 1976. At that time few details about test administration and scoring were readily available. Most of the published work on the DMT appeared in Scandinavian journals, written in Swedish, consequently the testing procedure used by the Royal Swedish Air Force could not be followed precisely. Nevertheless, in the first RAF trial, 82 pilots were tested as they approached the end of their Basic Flying Training (BFT) course. Sixteen of these students - 20% - subsequently failed BFT and 7 of these had been classified, by the DMT, as likely to fail. Unfortunately, by administering the test near the end of the BFT course, the sample did not include a substantial number of students who had been suspended from training earlier on during the 9 month flying course. Consequently it was considered that a more representative sample of student pilots should be obtained by administering the DMT prior to the start of basic flying training.

In June 1977 Kragh visited the UK. He suggested a number of improvements to the RAF DMT procedures and he offered the RAF his stimulus pictures. The information gained from Kragh provided the basis for the RAF's first proper DMT validation trial. Of course, it was realised at the time that an exact replication of the Swedish military work could not be carried out. Full details of DMT administration, coding and scoring were still not known by RAF aviation psychologists. Moreover, in Sweden the DMT was administered individually but it was considered to be impossible for individual testing to be carried out on an RAF sample since the number of RAF pilot applicants to be processed far exceeded the number tested in Sweden. However, Kragh had experience of administering the DMT to groups of subjects, so it was decided that group DMT administration would be suitable for the RAF.

From October 1977 to April 1978, 72 Direct Entrant pilot cadets were given a group version of the DMT towards the end of their 18 weeks Initial Officer Training course. Up to 6 cadets were tested together. They were shown distractor pictures and two sets of 16 threat pictures that were exposed in increasing durations ranging from 10 to 500 milliseconds. The subjects then had 2 minutes per exposure to draw a picture and write down what they saw. The test protocols were later scored by two psychologists using a coding manual written by Thomas Neuman in 1978.

The validity of the test was measured against two criteria:

a. Success at BFT, ie completing the 9 months BFT course and continuing fast jet training versus failure at this stage.

b. Success at the Tactical Weapons Unit (TWU) or failure at or before this stage.

Fifty students, out of those tested, entered BFT of whom 13 (26%) failed. There appeared to be a slight tendency for students with poorer DMT scores to fail BFT but the point biserial correlation of $r = .19$ was not statistically significant. By the end of TWU 31 (62%) had failed. Again the relationship between the DMT scores and training success was in the desired direction but the correlation ($r = .26$) was not significant.

In order to compare students' DMT scores with successful pilots, 35 operational fast jet squadron pilots were also tested. Their mean score was compared with the BFT failure group but yet again there was no significant difference.

The results from the first validation study were inconclusive but there was considerable interest in the project. Therefore a second validation was carried out between October 1978 and January 1980. This time the DMT was administered to an applicant population rather than a selected population - pilot candidates were tested at the Officers and Aircrew selection Centre at the end of the selection procedure.

By July 1982 criterion information was available for 158 subjects; 88 (54%) had failed BFT whilst 70 (46%) had passed. Unfortunately the results showed no relationship between DMT scores and BFT outcome. A similar finding was obtained when TWU results were examined (13 passes, 107 suspended earlier).

The two main RAF DMT studies had failed to corroborate the findings of the Royal Swedish Air Force (2). Group administration of the DMT was unsuccessful so further research using DMT in this manner was stopped. However it was acknowledged that there were methodological inadequacies in the RAF trials. The Scandanavian Air Forces were still producing remarkable results: the DMT was reducing training wastage and it was claimed to be playing a major role in reducing accident rates (3). Therefore it was considered possible that individual test administration might yield data that could predict RAF flying performance.

The opportunity to overcome previous DMT methodological problems was then presented when Thomas Neuman offered to organize a DMT trial. The trial began in 1984 and it is currently still being validated.

Initially Neuman trained four people to administer the DMT. They administered it to 253 RAF pilot trainees who had just completed Initial Officer Training (IOT) and who were about to begin RAF flying training. Testing took place during one year between April 1984 and May 1985.

All testing materials were loaned under contract to the Ministry of Defence by "Interpersona Ltd". Two holontachistoscopes were used to present the stimuli. These comprised sets of two distractors (depicting aggressive and sexual themes) and two percept-genetic threat pictures. The distractors were both shown before and after the two series of test pictures. The test pictures were each presented for 19 successive occasions, the exposure durations increasing exponentially from 5 to 870 milliseconds. After each exposure the subject drew a picture and the experimenter recorded the oral response. The subject was also instructed to guess the sex, age and mood of any person that he saw.

The test protocols were given code numbers by the testers and anonymous scripts were sent to Thomas Neuman for coding, scoring and interpreting. He then provided test scores and performance predictions for each of the 253 subjects. However, the exact method of test scoring and interpretation were not revealed to the Ministry of Defence under the contract. In order to ensure the objectivity of this trial, Neuman made predictions about the students flying performance based solely on the DMT protocols and these scores and predictions have not been revealed to flying trainers.

Two sets of DMT predictions were provided by Thomas Neuman before he was given any RAF flying training data. DMT results were based on his "Ten Aspects" scoring system from which the Neuman Pilot Index (NPI) was initially derived. This index is used to predict the ability to cope with basic pilot training and adaptability to operational flying (4). However Neuman was informed that the NPI did not correlate with Basic Flying Training performance therefore he issued a second set of predictions. The revised scoring system resulted from work that had been carried out on the development of a DMT system for predicting flying performance in the Royal Australian Air Force. From this analysis the Pilot Capacity Index (PCI) was dervied.

Point biserial correlations were calculated between the Neuman scores and flying training results. For the purposes of these analyses a pass at Basic Flying Training (BFT), Advanced Flying Training (AFT) or at the Tactical Weapons Unit (TWU) was recorded if the student was continuing in the fast jet training programme. Whereas a failure was recorded if the student were no longer continuing fast jet flying training (ie fail for airwork or other reasons, and streaming to rotary wing or multi-engine training were considered as a failure).

One hundred and fifty four students (61%) passed their initial basic flying training course and continued fast jet training. Fifty seven students (22%) were streamed into multi-engine and rotary wing training and the remaining 42 students (17%) failed flying training (35 of these (14%) were airwork failures). (See Table 1).

The original DMT scores (NPI) for this sample of 253 students ranged from 13 to 48 and formed a normal distribution. It was predicted that students with low DMT scores would be more likely to fail basic training and that those with high scores would pass. Unfortunately the training results did not support this prediction. The distributions of BFT fast jet passes, failures and multi-engine and rotary wing trainees were all spread across the DMT score range and the overall point biserial correlation between the DMT scores (NPI) and pass/fail BFT was non-significant, $r = 0.01$. The revised DMT scores (PCI) also showed a non-significant relationship with BFT results, $r = 0.07$. Moreover, further analyses revealed that none of the ten aspect component measures correlated significantly with BFT results. (See Table 2).

Some of the students who took part in this trial are still under training. However, at the time of analysis results were available for 229 students who had completed the Tactical Weapons Unit course or who had failed at, or prior to, that stage. The results so far show no significant relationship between the DMT scores and TWU training results.

It is concluded that the DMT scores do not predict success in RAF flying training. However the present DMT trial is not yet complete because Neuman also made accident and psychosomatic illness predictions. Neuman identified 17 students (7%) as being extremely accident prone (likely to kill themselves), 37 (15%) as highly accident prone and 53 (21%) as accident prone. In addition he predicted that 18 (7%) student pilots are highly likely to have psychosomatic illness and 26 (10%) are moderately likely to be ill. (The accident and illness groups are not mutually exclusive). Therefore accident and illness records are being monitored.

By the end of July 1988, accident records revealed that only one of the students in this sample had been involved in an accident where aircrew error was a possibility. (This student was not classified as being accident or illness prone by the DMT). Since 1981 the accident rate has been almost halved and in 1987 it was the lowest in the RAF's history (fewer than one major accident for 35,000 hours flown) (5). If this trend continues it will be unlikely that any significant relationship between DMT predictions and aircrew accidents will be established.

The RAF results with the DMT do not corroborate the Scandinavian findings. The reason why there is a discrepancy is not clear. It was believed that methodological differences had been eliminated in the present trial but this may not be the case. Where the DMT is used for military pilot selection, it is usually administered by clinical psychologists and the test forms part of a clinical interview. Therefore the interview information may be a vital integral part of the procedure.

Other explanations for the lack of generalization between the Scandinavian and the British results include differences in: the selection procedures and the selection ratios in Sweden and the UK, the flying training methods and operational flying performance, the way in which accidents are classified, and culture. However none of these explanations can be identified as the prime reason for the discrepant results.

The decision to employ the DMT as part of the RAF pilot selection procedure is a pragmatic one. So far there is no direct empirical evidence that suggests that the DMT could produce savings in costs, or life. Consequently there are no plans to introduce the DMT into the RAF selection procedure.

## REFERENCES

1. Neuman, T. Perceptual defence organisation as a predictor of the pilot's adaptive behaviour in military flying. 1971. Paper presented at the 9th WEAAP conference, Cambridge, UK.

2. Stoker, P. J. The validity of the Swedish Defence Mechanism Test as a measure for the selection of Royal Air Force pilots. 1982. CS(RAF) Note for the record 16/82.

3. Neuman, T. Influence of DMT on economy of training and flight safety. 1982. Interpersona Report.

4. Neuman, T. Dimensioning and validation of percept-genetic defence mechanisms. A hierarchical analysis of the behaviour of pilots under stress. 1978. FOA report C55020-H6. Translated from Swedish by R. E. Williams, 1985. HMSO London.

5. Statement on the Defence Estimates 1988. Volume 1. London, Her Majesty's Stationery Office. p. 29, para. 318.

TABLE 1 - BASIC FLYING TRAINING RESULTS

| RESULTS | NUMBER | PERCENTAGES |
|---|---|---|
| Pass fast jet | 154 | 60.9% |
| Multi-engine | 24 | 9.5% |
| Rotary wing | 33 | 13.0% |
| Fail air work | 35 | 13.8% |
| Fail medical | 4 | 1.6% |
| Fail other reasons | 2 | 0.8% |
| Fail voluntary | 1 | 0.4% |
| Total | 253 | 100.0% |

TABLE 2 - POINT BISERIAL CORRELATIONS BETWEEN DMT PREDICTIONS AND FLYING TRAINING RESULTS

| | Flying training results | | |
|---|---|---|---|
| | BFT(N=253) | AFT(N=250) | TWU(N=229) |
| Predictors | | | |
| NPI | .013 | -.012 | -.017 |
| C | .067 | .038 | .054 |
| HOLON | -.052 | -.021 | -.042 |
| U | -.010 | -.022 | -.068 |
| CORR | .009 | -.034 | -.041 |
| PATH | .012 | -.017 | -.029 |
| NC | .068 | .061 | .068 |
| ASEX | .089 | .040 | .048 |
| NVAR | .004 | -.000 | -.021 |
| T | -.005 | -.065 | .005 |
| BALANCE | .027 | .034 | .021 |
| NP3 | .020 | -0.13 | -.014 |
| DISTRACTOR | -.085 | -.139 * | -.074 |
| Revised predictors | | | |
| DEFENCES | .059 | .074 | .080 |
| DEFENCES N | .023 | .042 | .072 |
| POSITIVE WT | -.009 | .036 | .089 |
| NEGATIVE WT | .061 | .033 | .041 |
| AUSTRALIAN U | -.037 | .015 | .060 |
| PCI | .071 | -.013 | -.052 |

* $p < .05$
BFT Basic flying training
AFT Advanced flying training
TWU Tactical Weapons Unit

NB Details of the DMT scoring procedures and the derivation of the predictor measures were not provided under the contract between Interpersona Ltd and the Ministry of Defence.

## DISCUSSION

POLLACK: We have used the DMT for nearly 20 years in the Swedish Air Force. I think it is very important to know what lies behind these figures when we talk about DMT, and I think you should know more about our selection procedures. We have a step by step selection procedure which is, in fact, very much related to clinical personality work performed by our aviation psychologists who work in a selection centre all the year round. They work in a centre where we have squadron pilots who are well trained in selection techniques. When they have made a holistic assessment of the suitability of a candidate for the Air Force and the candidate is accepted, he then does the DMT. I would like to emphasise that the hard work of selection is already done before the DMT is performed. It is just about 2% of candidates who will be confronted with the DMT. When you read about the DMT you might believe that it determines who is selected for pilot training in Sweden. You cannot just take this test separately.

In fact, we have not done any evaluation of DMT for the Swedish Air Force. Thomas Nauman has done it, but we have not. So we cannot say if this test benefits our selection system or not. Our failure rate has decreased very much, but it is a result of us changing the selection procedure. We now take the cream of the milk instead of putting everyone into the training programme; it is very much related to the philosophy of our training today. You know we have changed aeroplanes during this period and everything else has changed, so you cannot rely on one single test.

I am an investigator working for the flight safety now. If you look at the incidence of our human factor related accidents the rate has not decreased at all since we started DM testing all pilots in the Swedish Air Force.

SIOMOPOULOS: I would like to offer an explanation as to why the DMT did not work with the RAF. I think the explanation is simple. When a clinical psychologist spends almost two hours with a candidate, surely you are going to obtain useful information. I do not know any system that has the luxury of having a clinical psychologist spending almost two hours with each candidate. Let's ask people from other countries. How much time does a clinically orientated person spend with its candidate? If they say almost two hours then you can really do a lot of things in two hours. If you want to make a trial, then try some people on DMT and some other candidates with just an hour long interview, the classical semi-structured interview, and see if there is any difference. I do not think there is going to be any difference because your system is very good and it could not be better. You do enough in the area of interviews. You do enough in all areas of your system, so it would not be improved by the introduction of other tests at interview.

WALKER-SMITH: For RAF pilot selection we use a 45 minute interview, which is *highly* structured, and this interview does contribute to the prediction of success of basic flying training. One possible explanation for the DMT not working is that when you tease out DMT data from interview data it is the interview that is predictive and not the DMT.

BOER: My Institute has some knowledge of the selection system of the Royal Netherlands Air Force. Perhaps there are others present here who can add to what I am saying, but what I know about the role of DMT in our selection it is an exact confirmation of what you have said. It was tried in our Air Force. A lot of time was spent on it. It took one hour per candidate, and we found out that we could not identify anything relating to success in pilot training that could be predicted from the test. So the decision has been taken to remove it from pilot selection.

BYRDORF: Just to add to what Dr Pollack from Sweden said. In the Royal Danish Air Force we use the DMT after our selection procedure. We have a step by step selection too and we have an interview by a clinical psychologist. The DMT is only a test put on top of the other tests in our battery and is used as a little bit more help. Sometimes we use it and sometimes we do not.

# LIGNES DIRECTRICES PRINCIPALES FONDANT LA SELECTION PSYCHOLOGIQUE DES CANDIDATS PILOTES A LA FORCE AERIENNE BELGE

J-C GENON * and P.VANDENBOSCH **

* Cdt J-C GENON , CHIEF PSYCHOLOGIST AT THE BELGIAN MEDICAL AEROSPACE CENTER

** Col P. VANDENBOSCH, MD., DIRECTOR AEROMEDICAL SERVICES, BAF.

KING ALBERT I QUARTERS, AIR FORCE BARRACKS, RUE DE LA FUSEE 70, B-1130 - BRUSSELS, BELGIUM

Il n'entre pas dans le propos de la présente réflexion de décrire les différentes épreuves utilisées pratiquement pour sélectionner les candidats pilotes de la Force Aérienne Belge.

En fait, il a été considéré que les techniques employées et leurs résultats sont souvent influencés, par des situations de fait de la vie courante ainsi que par des orientations propres aux praticiens en place. Parfois, ce sont là des éléments qui existent simplement en filigranne mais qui peuvent également prendre le pas sur les aspects scientifiques purs.

En partant de l'expérience de sélection auprès de la Force Aérienne Belge, il va être tenté de les analyser pour clarifier leur incidence sur les processus et les succès de la sélection.

## 1. Les fondements

### 1.1. Les préalables à la sélection

De façon générale, le recrutement puis la sélection ne sont pas loin d'une opération de marketing. Un organisme offre des fonctions à exercer dans le cadre de certaines conditions. L'attrait du produit c'est-à-dire la fonction intéressé va déterminer la demande, autrement dit le nombre et la qualité des candidatures. Ce sont des éléments comme la rémunération, le statut administratif, la formation, le développement d'une carrière qui vont en grande partie influencer beaucoup de sujets à côté d'autres facteurs moins pragmatiques comme, dans le cas de la sélection aéronautique, le plaisir de voler. Comme, en général, toutes ces caractéristiques sont fixées pour longtemps, le sélectionneur devra bien vivre avec la nature de la population qu'elles vont drainer.

De la même façon, le volume ainsi que le contenu et la forme de la publicité influencent sensiblement la population à sélectionner.

Appliquées à la Force aérienne belge ces considérations expliquent deux mouvements actuels :

- des efforts de l'autorité pour revaloriser la fonction de pilote de combat.
- des investissements pour rehausser la qualité et la précision de notre publicité.

D'autre part, certains faits de société compliquent encore la constitution d'une population intéressante. Une désaffection pour les carrières militaires et pour les problèmes de défense écarte nombre de candidats potentiels. D'une certaine façon, dans un état de droit, avec une police organisée, lorsqu'aucune mobilisation n'existe envers des menaces extérieures éventuelles, la fonction psycho-physiologique de défense s'émousse dans la société et, en une sorte de formation réactionnelle, les expressions pacifistes augmentent.

Un autre fait de société concerne l'évolution de l'enseignement. Les niveaux d'exigence tendent à baisser et les jeunes gens perdent l'expérience de la lutte contre les barrières comme les classements et les examens. A maintenir nos propres exigences par des standards de sélection élevés, nous avons déjà couru le risque de nous retrouver sans candidats admis. De là, il apparaît une tentation de baisser nos seuils, ce qui se révèle défavorable à long terme, mais aussi une nécessité de compenser certaines lacunes de l'enseignement en investissant davantage du temps et de l'argent dans les écoles des Forces armées.

A ces aspects, s'ajoute le problème de l'offre et de la demande. La situation économique étant ce qu'elle est et le déficit en pilotes de combat étant notable, beaucoup de candidats se présentent pour un petit nombre de places ouvertes. D'où des problèmes pour les examiner. Il faut pouvoir les absorber tous dans des délais raisonnables ce qui va à l'encontre d'une sélection en profondeur.

Les procédure et les techniques de sélection appliquées auparavant ne suffisent plus et une adaptation s'est révélée nécessaire, par exemple, par un effort dans l'informatisation et par l'accentuation des différents palliers de sélection.

### 1.2. Procédures de sélection

Les conditions d'admission comme l'âge, ou le niveau d'étude, n'ont pas été changées. Le nombre de candidatures n'est donc pas diminué à ce niveau, au contraire, il risque d'augmenter quelque peu puisque la fonction de pilote de combat est ouverte maintenant aux jeunes filles. Aussi, a-t-il fallu soigneusement répartir la succession des épreuves en fonction des critères de coût, de sélectivité et de durée. Dans la pratique, lors d'une première convocation, les organes des sens sont d'abord investigués puis les aptitudes psychomotrices. A ce moment, environ 60 % des candidats sont éliminés et lors d'une seconde convocation, la personnalité, les capacités sportives et les autres domaines médicaux sont abordés.
A l'heure actuelle, les épreuves académiques demandent une troisième convocation mais, vu leur pouvoir d'élimination et de classement, elle seront déplacées dans l'avenir au premier stade de la sélection juste après l'examen des organes des sens.

### 1.3. Chaîne de sélection

Il y a une dizaine d'années, la sélection constituait une étape totalement isolée du recrutement et de la formation. Ce qui ne pouvait entraîner que des faiblesses. Une évolution progressive, bien soutenue par les autorités de la Force aérienne, a conduit au concept nécessaire d'un continuum "recrutement - sélection - formation".

En particulier, chaque élève radié est réexaminé par le psychologue examinateur. Au-delà des travaux chiffrés de validation, les failles de la sélection et de la formation sont ainsi bien mises en évidence et, dans la mesure du possible, des corrections sont apportées.

L'idéal serait de réaliser un "suivi" des élèves au moins jusqu'à l'obtention des ailes mais l'appareil d'appui psychologique ne le permet pas et il faut se contenter d'interventions ponctuelles sur demande des responsables des écoles ou des médecins des unités. La situation n'est pas satisfaisante à l'heure actuelle, d'abord vu le nombre de psychologues en place, mais aussi suite à l'image du psychologue auprès du personnel navigant. C'est un problème de présence et de contact qui exigent beaucoup de temps et dont on peut penser qu'il ne sera jamais tout à fait résolu. Pour des raisons de rendement, il nous est difficile d'intervenir avec efficacité auprès des élèves. D'autre part, des problèmes dans l'apprentissage au vol n'entraînent pas nécessairement une demande d'aide de la part des élèves et lorsqu'un cas connaît une évolution heureuse, l'intéressé n'en fait pas de publicité ce qui ne fait pas évoluer la réputation de l'aide psychologique.

1.4. Responsabilité de commandement

En synthèse, il nous semble important de souligner que la qualité d'une sélection est autant le résultat d'une politique de commandement que de l'action des spécialistes qui l'éxécutent. Quand un commandement s'intéresse à la sélection de son personnel, il prend davantage conscience des exigences qu'il pose et des moyens nécessaires pour y répondre. En particulier, c'est à lui, en tant que tiers commettant, de fixer les seuils d'acceptation après consultation de ses spécialistes. C'est à lui aussi qu'il appartient, en dernier ressort, de maintenir l'équilibre entre les moyens déployés, le volume d'examens à pratiquer et les résultats qualitatifs qui vont en découler.

## 2. L'approche du candidat

Deux positions importantes, lors de l'examen des candidats, demandent un effort permanent afin d'effectuer un travail de sélection efficace.

2.1. L'équipe de travail

En premier lieu, il convient de réaliser une collaboration interdisciplinaire la plus confiante possible. Ce n'est pas une grande nouveauté qu'une telle affirmation mais il est probable que l'énoncé en est plus aisé que la réalisation dans la pratique quotidienne. Chaque spécialiste oeuvre dans son domaine propre principalement mais, sans la connaissance des autres approches ou des autres points de vue, l'évaluation d'une candidature peut rester incomplète. Par exemple, une tension élevée peut conduire au rejet d'un candidat lorsqu'elle va se conjuguer avec une défense inadéquate en situation de stress. Ou encore, la connaissance aéronautique vécue d'un instructeur pilote peut enrichir l'approche clinique ou testologique du psychologue.
Enfin l'adjonction directe de statisticiens - informaticiens, bien documentés quant aux demandes, auprès des médecins et des psychologues élargit les demandes, auprès des médecins et des psychologues élargit les recherches et les études de validation. Déjà résolu, dans beaucoup de pays, ce problème vient d'être abordé à la Force aérienne belge par la création d'une section informatique au Centre Médical Aérospace.

2.2. L'élève potentiel

Les candidats sont des sujets jeunes, encore liés au milieu parental pour la plupart, mais au seuil de leur vie professionnelle et de leur vie familiale. Fort peu d'entre eux possèdent une image réaliste de l'avenir dans lequel va les plonger leur acte de candidature, en particulier, la situation de combat aérien. Plus qu'une motivation, terme trop popularisé et vidé de son sens, c'est envers l'évaluation d'une capacité d'adaptation et d'intégration qu'il y a lieu d'orienter l'appréhension de leur psychisme. Ce n'est pas là une position simple qui va exiger l'établissement d'un pronostic, donc d'une projection sur l'avenir dont il sera question plus loin. En fait, on pourrait dire qu'un suivi complet débute à ce moment, compte tenu de cette relation privilégiée que sera la relation "élève - moniteur". Bien sûr, le moniteur est encore inconnu mais, si la collaboration psychologue - moniteur était bien développée (dans le sens où chacun peut apprendre de l'autre pour le meilleur intérêt des candidats et des élèves) la base d'un suivi dynamique serait jetée lors de l'approche psychologique du candidat.

## 3. La sélection psychologique

3.1. L'acte psychologique

Que se soit lors de l'établissement initial d'une procédure de sélection ou durant la sélection quotidienne elle-même, deux attitudes devraient rester présentes en permanence à l'esprit : l'individualisation des examens et l'acte psychologique en tant qu'acte relationnel.

Cela ne veut pas dire que l'entièreté des épreuves doit être administrée individuellement. C'est une impossibilité de fait vu le nombre des candidats. Au contraire, la plupart du temps, des filtres successifs doivent être posés pour éliminer les sujets les plus faibles surtout dans le cadre d'une sélection quasi "tout - venant".

Mais il est souhaitable que la dernière partie de la sélection soit plus individualisée lors d'un entretien et/ou lors d'administration d'épreuves de personnalité. Des objections pratiques ou financières sont souvent présentées à l'encontre de ce mode de faire. Sans minimiser l'obstacle, il peut y être répondu que les investissements en nombre et en formation du personnel occupé de psychologie sont rentables à long terme.

L'accent porté sur l'acte relationnel de l'examen est une conséquence de l'individualisation. Un premier aspect porte tout simplement sur la qualité humaine de l'accueil. Un second consiste à concilier, dans le cadre de l'examen et des conclusions mêmes, les intérêts du candidat et du tiers commettant c'est-à-dire la Force aérienne. Paradoxalement, ces intérêts se recouvrent : la Force aérienne veut miniminaliser les taux de radiation, et il est d'importance pour les candidats de ne pas s'engager dans une voie sans issue. Enfin, une sélection à l'aveugle, l'emploi du psychologue par lui-même comme instrument d'évaluation (autrement dit un sens clinique fondé sur les réactions affectives propres de l'examinateur) est impossible. Or, il est de notre politique d'examen de privilégier la synthèse clinique dans l'approche du psychisme.

3.2. L'interprétation

Cette synthèse clinique vise quatre buts :

- effacer le clivage entre ce qui est "aptitudes" et "personnalité".
- éviter une "atomisation" des sujets par toute une série d'éléments de personnalité investigués séparément mais en interaction réelle dans le fonctionnement psychique.
- tenter de nuancer la différence entre les situations d'examen et de réalité (c'est-à-dire en écolage ou en vol).
- évaluer un potentiel susceptible d'évolution plutôt que brosser un portrait statique.

3.3. Les critères

Mais l'interprétation conduit inéluctablement à la décision finale "acceptation - rejet - rejet nuancé" de la candidature.

Les seuils d'exclusion fondés sur des recherches de validation statique sont bien connus et utilisés universellement. Une seule remarque à formuler est que peu de foi est accordée à la recherche de validité sur des critères apportés par une épreuve isolée. Cette approche est nécessaire pour améliorer l'outil mais l'expérience montre peu de résultats encourageants. Encore une fois, le "sujet vu comme un tout" et gratifié d'un pronostic de réussite constitue la pierre angulaire de la critique de notre sélection.

Une autre base de jugement se rattache au concept de "sélection négative". Il n'est pas facile (et, probablement, n'est-ce pas possible) de définir ce qui fait qu'une personne obtient un jour un brevet supérieur de pilote. Certaines caractéristiques comme la confiance en soi - ont été pressenties mais aucun modèle cohérent et complet n'existe à ce jour.
La sélection négative, qui consiste à ne voir aucune contre-indication à une candidature, sans garantie de succès ou d'échec en écolage, peut être une position confortable pour l'examinateur à condition qu'il ne conclue pas sur un seul signe révélé par les épreuves mais bien à partir d'un faisceau de signes. Pourtant, s'en tenir aux seules contre-indications, ne nous semble que nécessaire mais insuffisant pour progresser dans la connaissance psychologique de candidat pilote.

En particulier, et ce n'est ni une boutade ni un commentaire péjoratif à l'égard des navigants, il n'est pas certain que l' "idéal de maturité psychique" qu'un psychologue puisse se représenter soit aussi l'idéal pour être un bon candidat. Autrement dit, il est probable que certains fonctionnements défensifs soient une aide ou une prédisposition pour une certaine activité qu'elle soit aéronautique ou autre. Il n'y a rien de neuf dans cette observation non plus mais en position de sélection, il faut se garder de ne pas classer comme cause de rejet tout ce qui n'est pas du domaine de la stricte intégration des problématiques dans la mesure où l'adaptation reste réalisée.

4. Les modèles de fonctionnement

Ce qui précède a peut-être pour conséquence de donner l'impression que la sélection fondée sur le fonctionnement psychique reste incertaine ou hésitante.

Une telle impression serait inexacte puisque la psychologie expérimentale et la psychologie clinique ont, à l'heure actuelle, mis en évidence beaucoup d'éléments de la réalité psychique.
Pourtant, la connaissance reste bien incomplète, surtout par manque d'un modèle unique du fonctionnement neuro-psychologique. Plusieurs modèles fort valables existent qui sont utilisables dans la pratique de sélection courante. Néanmoins, nous avons souvent eu le sentiment que chacun privilégie une partie des phénomènes psychiques et que des fonctionnements identiques sont décrits avec des terminologies différentes sans que les auteurs reconnaissent ces identités.

Une position plus éclectique peut consister à rassembler ces modèles en un seul mais l'incomplétude de nos connaissance ne rend pas cette tentative totalement possible.

En essayant de conserver une attitude ouverte, nous avons essayé de minimiser une approche fondée sur les traits de personnalité seuls, conception trop statique et trop atomistique pour privilégier un modèle de référence plus dynamique où les mécanismes de défense et d'adaptation prévalent au même titre que l'interaction dynamique entre le sujet et ses différents milieux.

## 5. Les développements ultérieurs

### 5.1. La stimulation

Dans le domaine des techniques d'investigation, l'informatisation a permis beaucoup de progrès dans la simulation des situations réelles. Encore ne faut-il pas confondre simulation et sophistication.

Il n'y a pas lieu de discuter ici l'intérêt du remplacement des épreuves "papier - crayon" par une présentation informatisée pour obtenir un gain de temps lors des corrections et des exploitations statistiques. Le gain est évident. Il faut toutefois prendre garde aux modifications de la relation "examinateur" - "candidat" que ce progrès implique.

De même, l'intérêt de nouvelles complexifications des épreuves proposées nous semble évident. On sait depuis longtemps que, pour des fonctions de haut niveau, c'est dans le cadre des épreuves les plus complexes que se produit la meilleure différenciation entre les sujets.

Par contre, nous prenons nos distances vis-à-vis d'une sophistication exagérée des appareils de sélection dans le domaine de la psychomotricité.

Un appareillage très complexe, proche d'un simulateur de vol, appartient au domaine de la formation et non de la sélection. Pour celle-ci, l'existence de tâches séparées puis simultanées mettant en oeuvre les mêmes fonctions exigées dans le vol réel suffit même si ces tâches ont la forme de situations de laboratoire.

D'autre part, en ce qui concerne le vécu émotionnel des candidats, nous pensons que la simulation dans la sélection reste toujours limitée, vu les différences de signification affective entre les situations d'examen et de vol réel. Pour cette raison, notre sélection octroie une place à l'investigation par des stimulations astructurées que nous pensons plus susceptibles de mettre en évidence les mécanismes de défense et d'adaptation ainsi que de percuter les points sensibles des personnalités.

Enfin, la popularisation et le développement des jeux informatiques proches des épreuves de psychomotricité risquent peut-être d'introduire des distorsions dans l'évaluation des candidats. Sans moyens pour expérimenter ce danger nous nous sommes référés à des travaux étrangers mais ceux-ci sont parfois contradictoires.

Quoi qu'il en soit, le choix d'épreuves pour une future batterie informatisée nous amène à une certaine circonspection en essayant d'introduire des épreuves plus éloignées des appareils commercialisés. De toute façon, les générations futures ne développent-elles pas des 'engrammes ou des métacircuits différents suite au développement de la stimulation par l'image et de la pratique courante et précoce de claviers de commande ?

## 6. La prospective

Cette petite incursion dans l'anticipation nous amène à aborder nos orientations et nos intérêts pour l'avenir.

Un premier point, se situe dans le prolongement de nos références, actuelles.
Des épreuves nouvelles fondées sur l'interaction entre le sujet et son milieu, ainsi que sur la complexité des situations d'examen devraient compléter la testologie traditionnelle. Malheureusement ces techniques sont encore au stade du balbutiement.

Une autre direction possible, pour améliorer l'investigation classique de la personnalité est l'infraperception. Un programme est en cours afin de réaliser une première recherche.

Il reste enfin les examens psycho-physiologiques soit dans le domaine de l'électricité cérébrale soit dans le domaine des catécholamines. Compte tenu de la précarité de nos moyens, il ne nous est pas possible de nous engager dans ces voies de recherche sauf pour des incursions très restreintes et tout-à-fait exceptionnelles mais, étant d'opinion que la coopération entre le physiologiste et le psychologue constitue une voie royale pour l'avenir, nous tenterons de suivre les évolutions de près.

# PREDICTION OF SUCCESS IN FLIGHT TRAINING BY SINGLE- AND DUAL-TASK PERFORMANCE

by

P.G.A.M. Jorna
TNO Institute for Perception
Kampweg 5,
P.O. Box 23
37692G Soesterberg
The Netherlands

**Summary**

Advanced technology has changed the type and the amount of information a pilot has to process. The military pilot is particularly involved in performing multiple tasks under difficult working conditions. Some aspirant pilots are not able to finish advanced training, apparently for reasons of an insufficient multiple task performance. A reduction of such attrition is highly desirable to reduce the cost of training. A test based on dual-task performance has been developed to investigate the trainability of aspirant pilots to perform under such demanding conditions. The dual-task was a combination of a pursuit tracking task with preview and a continuous memory task (CMT). Aspirant pilots practiced the tracking task and were tested under single- and dual-task conditions. Dual-task performance was expected to be related to pilot aptitude as assessed by other criteria.

The results were validated against the level of pilot aptitude as assessed by traditional selection procedures, a flight simulator test and advanced flight training for the Lockheed Orion and the Westland Lynx helicopter. Successful aspirants, now operational pilots, were characterized by their excellent performance under dual-task conditions. Less successful aspirants performed less efficient under dual-task conditions as well as single- task conditions, depending on how soon they failed in the selection and training process. Prior flying experience did not influence tracking performance and was not found to be a critical factor in predicting success in advanced or operational flight training.

## Introduction

Pilot selection is very strict and only a small percentage of the applicants is allowed to enter flight training. Despite, a substantial number of those who enter basic training do not complete advanced training or acquire an operational status. Military aviation requires the pilot to integrate "flying with fighting", which involves controlling the aircraft while simultaneously managing complex systems. Performance has to be maintained even under the stressful and dangerous working conditions that characterize military aviation. The Royal Netherlands Navy found that aviators, who apparently had no problems with flying the Fokker Friendship, did encounter difficulties in handling the Lockheed Orion under operational circumstances. According to the instructors' view, these difficulties occurred in particular when parts of the mission required task integration or the handling of multiple tasks. The traditional selection procedures apparently did not cover some of the abilities or skills required in handling this modern aircraft with its complex technology. Time-sharing or dividing attention in multiple task situations seemed to be relevant for success in advanced flight training. A reduction of individual differences in multiple task performance by selection, could reduce the attrition in this part of the training program. This study will report on the development and validation of such a selection test.

The study was initiated in 1981 and there were two theoretical positions that could serve as a guideline for test development. The first line of research tried to identify a specific time-sharing factor, either an ability or skill, that is only relevant in dual-task situations. The second one, however, explained variations in dual-task performance by differences in the effectiveness of information processing under single task conditions. We therefore investigated the predictive validity of both single- and dual-task measures. The final test included training conditions in order to assess the "trainability" of a pilot candidate to perform multiple tasks relevant for the aviation environment. All subjects were tested prior to the normal selection and training procedure. Their career was subsequently followed to investigate the potential of single- and dual-task measures in predicting which candidate would pass or fail a certain step in the selection and training process.

## Theoretical Considerations

The first theoretical notion explained individual differences in multiple-task performance by postulating the existence of a specific time-sharing ability. Dual-task situations should provide information about performance efficiency that cannot be obtained by testing under single task conditions only. The quest for identifying a general factor, i.e an ability or skill relevant for performance in different task combinations, has not been very successful and the possibility of ever finding one is still under debate (1, 2). These results do not deny the importance of appropriate planning and allocating attention for complex performance, but indicate that there does not seem to be a unique skill important for performance in all kinds of multiple task situations. This implies that it is unlikely that any pair of tasks will be successful in predicting pilot performance. A first requirement for the tasks to be designed is that they should be relevant for flying or handling an aircraft.

The second theoretical notion states that dual-task performance is mainly limited by the capacity of human information processing resources. Performance will deteriorate only if these resources are exceeded. The relation between resource expenditure and level of performance is non-linear, dynamic and can be changed by practice as described by Norman and Bobrow (3) and Shiffrin and Schneider (4). Individual differences in dual-task performance are explained by differences in the relation between performance and resource expenditure at a single task level. Testing under single task conditions should then be predictive for performance under dual-task conditions. It is expected that single task performance will have predictive validity until performance cannot be improved substantially by trying "harder" or when it reaches an asymptotic level due to training. Performance is then thought to be data limited. This line of thinking resembles the practical observation that many subjects are able to perform under "normal" conditions, but that only a few perform satisfactory under stressful or more complex task conditions. Dual-task conditions serve the purpose of revealing a difference in "spare capacity" that is not reflected by differences in single task performance. An easy task will lose its "performance sensitivity" sooner then a more difficult task. Multiple resource theory as described by Navon and Gopher (5) and Wickens (6), however, showed that there are several resources that underlie human performance. Some task combinations will therefore produce more dual-task interference than others. The interference will be a function of the magnitude of overlap between specific resources needed for the execution of the tasks. Only when both tasks share (a) common resource(s), they produce interference. If no resources are shared, increasing the difficulty of one task will hardly affect the performance of the other. This is known as the "difficulty insensitivity" phenomenon.

Task structure is important but timing requirements could also influence the way in which resources are being shared by tasks. An overlap could be theoretically reduced by a rapid switching between tasks. In that case there is a common resource involved in information processing, but there is no divided attention or resource competition. In order to do so, both tasks should also require information processing at the same time. Otherwise, performance decrements could be found that are related to problems with switching and focussing attention instead of overloading a specific processing resource. Such switching or timing could be more optimally assessed in other experimental paradigms, such as the Dichotic Listening Test designed by Gopher (7). The use of a discrete "secondary task" with fixed interstimulus intervals should therefore be prohibited as it could induce a switching strategy. Tasks wich impose a more continuous demand should increase resource competition. It could be considered to use the same or highly similar tasks, but this would be problematic for reasons of sensory or effector interference, or the possibility of task integration in which case the dual-task condition will cease to exist. Similarities or a consistent timing relation between the tasks should be prevented as much as possible.

A final consideration is that resource competition is not only affected by task structure and timing, but also by the level of training. Well practiced subjects are assumed to require less processing resources for achieving a certain level of performance, which will be an advantage in dual-task conditions. Training should be provided in order to assess the trainability of a subject for a specific task. Additionally, it cannot be excluded that practice might be involved in the development of a specific time-sharing skill that is unique to the dual-task situation (8). Unfamiliar tasks should be trained under both single- and dual-task conditions in order to allow changes in the performance/resource relationship to occur as well as to reduce the "costs" of handling both tasks simultaneously.

**Task Design and Procedure**

The dual-task designed for this study, was a combination of a pursuit tracking task and a continuous memory task (CMT). The CMT (9, 10) is an auditory task administered by means of a prerecorded tape. The tape contains letters of the Dutch alphabet (consonants) presented at irregular intervals. As illustrated in Figure 1, the subject has to memorize a set of four letters denoted as the positive or target set. Target letters require a response by the non-dominant hand. The most demanding part of the CMT is that the number of targets has to be counted by keeping four separate tallies in working memory.

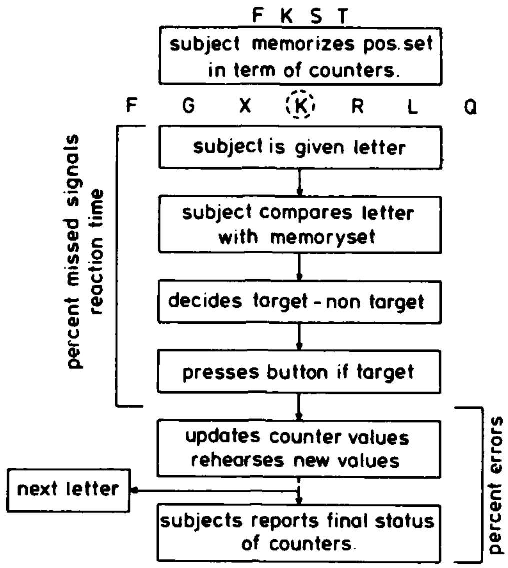



Figure 1. Structure of continuous memory task (CMT) with specific performance measures (integrated performance measure excluded).

The CMT has two important features: learning can be controlled by changing the letters in the target set. This will force the subject to maintain attention demanding "controlled processing" of information as described by Shiffrin and Schneider (4). Secondly, information processing is continuously required for the rehearsal and updating of the tallies, in addition to memory comparison, responding and counting. This task is so demanding that it is hardly possible to adopt a strategy of rapidly switching between the two tasks. Task load can be maintained at a high level even when the task is executed several times. An integrated performance score is obtained by calculating the discrepancy between the number counted and presented for each target. The sum then is related to the total number of targets and expressed as a percentage error. This score can be corrected for non-detected targets as indicated by a response omission. The measurement of reaction time provides a control for false alarms and late responses.

The pursuit tracking task provides preview, which enables the subject to progress from a simple error controller to one using preview for anticipatory actions (11). The display of the tracking task is illustrated in Figure 2. Pursuit and preview displays present more information than compensatory displays. Performance will only improve if the controller is able to use the additional information. The necessary control actions must be anticipated and stored in working memory for later execution. Such information processing is expected to be less efficient under dual-task conditions, in which simultaneous processing of other information is mandatory. As a consequence performance should deteriorate, especially if the other task is as difficult as the CMT. Such a decrement in quality of control should be minimal for aspirant pilots.

The subject has a direct control over the position of the window in stead of controlling the velocity or acceleration of that window. This simple control law was selected to prevent a possible advantage for subjects with prior flight experience. More complex types of control require extensive practice and pilots are better trained in such types of control.

The track itself was presented by a PDP 11-34 computer. A track contained a saw-tooth type of signal with rounded curves. The sequence and amplitude of the curves was randomized for each track. With this procedure it is possible to generate a number of different tracks with equal complexity. All the tracks are stored in separate computer files. Performance is scored by the Root Mean Square

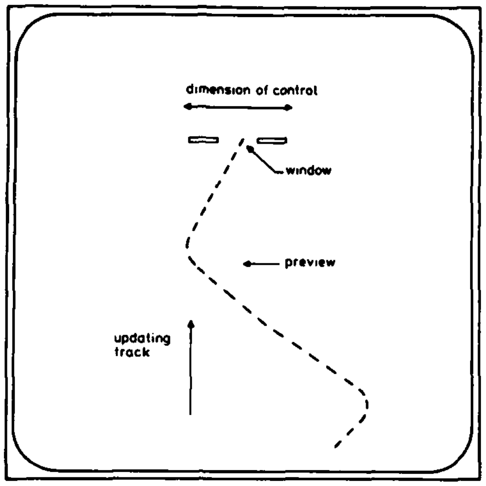



Figure 2. Unidimensional tracking with preview.

(RMS) error index. The track was designed in such a manner that it required the subject to continuously attend the track, in order to prevent it from leaving the window. Task interference with the CMT was expected to occur as both tasks require memory processes. It has been reported in several studies that tracking and memory tasks interfere consistently (6). Both tasks impose a continuous demand, in order to prevent a strategy of switching between tasks and a visual/auditory combination minimizes sensory interference. There is no consistent timing relation between tracking and the CMT and initial try-outs revealed that pressing the response button did not interfere with tracking.

After a familiarisation with tracking and a pre-test on the CMT, subjects performed in three task blocks. Each block contained single tracking, tracking with CMT and single tracking. The pre-test of the CMT and the results of block I provided performance scores for a situation with minimal practice. The results of block III and a post-test of the CMT were used to assess performance after additional training under both single- and dual-task conditions. The target letters were changed in the last block. The duration of each task condition was seven minutes. A rest period of five minutes was given between each test block. All subjects were instructed that they could improve their tracking by using the preview and that both tasks should be executed as well as possible in all conditions. Subjects were not aware of the fact that this test was not a part of the normal selection procedure.

## Validation Studies

The validation of a test procedure is difficult if there are no objective criterion measures. Pilot aptitude, however, is assessed at several steps in the selection and training procedure. Subjects that pass more or even all steps, are considered to have a higher level of pilot aptitude as compared with subjects that failed during this process. If dual-task performance is relevant for success, performance is expected to be better for those subjects that pass some, several or even all steps of this process. The selection and training procedure for naval pilots is summarized in Figure 3.

One way of validating the task is to test subjects after they passed or failed a certain stage in the training program. Such a comparison could provide evidence for the relevance of dual-task performance in such failure i.e concurrent validity, but does not provide a reliable indication of predictive validity. In order to do so, the test results should be obtained prior to entering flight training. At that time it has to be decided if pilot aptitude is high enough to assure a fair chance of finishing training all the way. The test was therefore validated by testing aspirants at our laboratory before entering the selection and training procedures. Their performance was subsequently related to the actual level of pilot aptitude as defined by the throughput of subjects at different steps of the selection and training system. This implies that there was a time period of up to six years between the experimental test and the final results of training.

The first step in the validation was to compare the test performance of high and low pilot aptitude subjects as defined by the outcome of traditional selection procedures. Subjects that did not pass, will be denoted as the "low

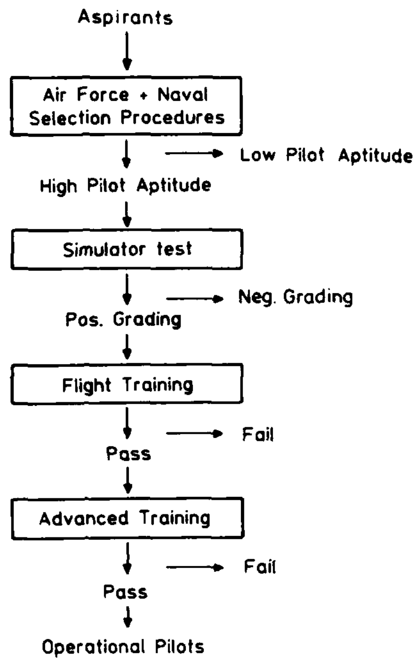



Figure 3. Selection and training procedure for naval pilots.

aptitude" group (n=14). Subjects that did pass will be denoted as either "pilot" (n=6) or "high aptitude" group (n=5), depending on having prior civil flying experience (A-type license).

The next step was to compare aspirant pilots that passed or failed a "flight grading" on a simulator. This grading is performed by experienced flight instructors of the government school for aviation in the Netherlands. If successful, by instructors grade, they are admitted to flight training. Subjects that passed will be denoted as the "positive grading" group (n=16) and those that didn't, as the "negative grading" group (n=12). All of these subjects had passed traditional selection and constituted a larger sample of "high aptitude" subjects.

The last and most important step was to contrast successful pilots with those who failed. Flight training was regarded to be successful only, if the pilot obtained an operational status on the Lockheed Orion or the Westland Lynx. Finishing basic training is sufficient to become a pilot, but advanced training is necessary to be allowed to fly missions with military aircraft. The group that passed the flight grading (n=16) also passed basic flight training and entered the advanced training program together with some certified pilots (n=4) who didn't receive a flight grading because of their extensive experience (B-type license). Pilots who became fully rated will be denoted as either the "Orion" (n=6) or "Lynx" (n=8) group and pilots who didn't finish training will be denoted as the "fail" (n=6) group.

The statistical contrasts were made by means of the Mann-Whitney U test, a non-parametric test based on ranking. This test is especially suited for our purposes because it provides information about the consistency of ranking one group as superior to the other. The test will provide absolute levels of significance as obtained with that specific ranking. A better level of significance (one-tailed) will indicate a better discrimination between the groups of interest.

## Traditional Selection

Aspirant pilots received the full battery of tests as used by the Royal Netherlands Air Force as well as the screening of the Royal Netherlands Navy. We now contrast the results of aspirants that passed both these procedures with the results of those who failed. The performances of the "high" and "low aptitude" group for tracking with preview under single- and dual-task conditions are summarized in the left and right panel of Figure 4.

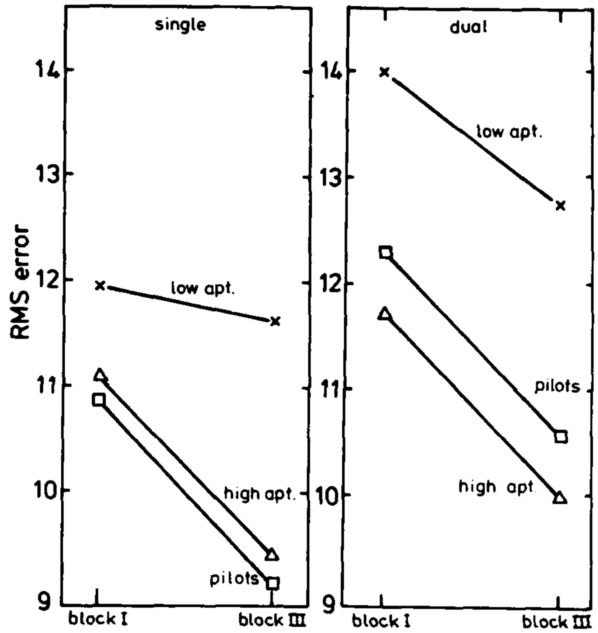



Figure 4. Tracking performance at the beginning (block I) and end (block III) of the training session for both single- and dual-task conditions.

Under single task conditions, the high aptitude and pilot group improved their tracking performance from block I to III, whereas the low aptitude group did not show an improvement as a function of training. As a consequence, both the pilot ($p<0.008$ Mann-Whitney U) and high aptitude group ($p<0.025$) obtained a better level of performance. Training on tracking resulted in a larger difference beween groups. The low aptitude group was not able to use the preview in an effective manner. An important observation is that pilots also needed training to improve their performance. This confirms that tracking performance was *not* confounded by prior flying experience.

Under dual-task conditions, the differences between the groups were larger as compared with the corresponding single task conditions. The separation of the groups was more consistent as indicated by a higher reliability or level of significance. Dual-task performance in block III, provided the best distinction between the low and high aptitude group ($p<0.004$). Note that pilots performed less than high aptitude subjects in this condition, whereas an opposite trend was present under single task conditions.

The results for the CMT are summarized in Figure 5. Again, the pilot and high aptitude group could be distinguished best under dual-task conditions. Especially the last block involving a new set of target letters, resulted in a better performance for the pilots ($p<0.001$) and high aptitude group ($p<0.001$) as compared with the low aptitude group. Similar differences were found for the other conditions. Even the pre-test indicated a distinction between the low aptitude and other subjects ($p<0.03$). Note that only under dual-task conditions the pilots performed less as compared with the high aptitude group. A similar pattern as observed with tracking.

The performance levels obtained for both tasks resulted in a similar ranking of groups differing in pilot aptitude. This supports the assumption that the tasks share a common factor or resource, thereby producing consistent dual-task interference. This interference was small or even not present for the high pilot aptitude subjects. The low aptitude subjects as identified by present selection procedures did indeed perform worse, especially under dual-task conditions. This provides support for the adequacy of the selection decisions. The pilot group, however, performed well under single task conditions, but worse under dual-task conditions. Being able to fly an aircraft does not seem to be a guarantee for adequate dual-task performance.

## Flight Simulator Testing

The high aptitude subjects were submitted to a Flight grading on a GAT 3 type of simulator, to assess their trainability. They received six sessions of about one hour, distributed over 3-4 days at which time they were guests of the aviation school. The performance evaluation resulted in either a positive or

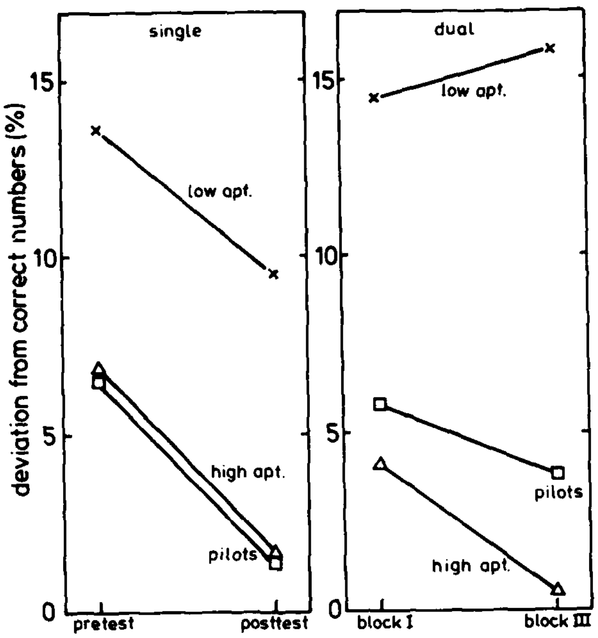



Figure 5. CMT performance in the pre- and post-test, and at the beginning (block I) and end (block III) of the training session.

negative grading as determined by two separate instructors. About half of the subjects did not pass this stage in the selection process.

Figure 6 shows a two dimensional plot depicting performance for both tasks under single- and dual-task conditions. Each axis provides a scale of measurement for a particular task. The performance level obtained under single task conditions is simply indicated at the appropriate scale. Together they provide a coordinate for dual-task performance as predicted in case of no interference. Such interference, however, is expected. By plotting the actual dual-task scores it is possible to inspect the relative dual-task decrements as well as to contrast groups on their absolute level of performance, as obtained under both single- and dual-task conditions. The data points were obtained in block III, which had the best distinction between groups.

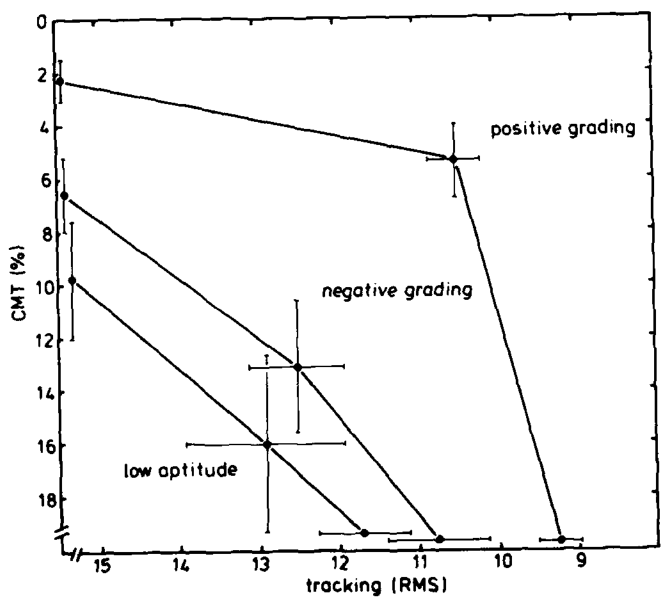



Figure 6. Single- and dual-task performance for aspirants with different pilot aptitude as assessed by selection ("low aptitude") and a flight simulator test ("positive/negative grading"). Standard errors indicated for both tasks.

The figure shows that differences in single task performance were still present between aspirant pilots who proved to be successful at this final step of selection, and those who didn't. The positive grading group performed better on both the tracking task ($p<0.025$) and CMT ($p<0.003$). They performed well under dual-task conditions. Both tracking ($p<0.002$) and the CMT ($p<0.002$) provided better scores as compared to the negative grading group. Dual-task conditions could serve to confirm and extend the conclusions that could be based on single task scores. The advantage is a better distinction as indicated by a higher level of significance.

The low aptitude group performed worse as compared to the negative grading group. This observation provided evidence for a consistent relation between the absolute level of performance as obtained by the test and the level of pilot aptitude as determined by selection. This was observed for both single- and dual-task performance. Additionaly, selection resulted in more homogenuous groups, as indicated by a smaller standard error for groups considered to have a higher pilot aptitude. The relative difference between the low aptitude and negative grading group was, however, rather small. This suggests that some of the subjects in the negative grading group should have been eliminated at an earlier step in the selection process. Pilot aptitude is possibly being overestimated by the traditional selection methods.

Both single- and dual-task conditions correlated significantly with the grades given by the instructors. These grades effectively represented a five point scale, with the lowest grade a 4 and the highest an 8. All values equal to or higher than 6 represent a "positive grading". Dual-task scores provided the best correlations. Dual-task tracking resulted in a Spearmann Rho correlation of 0.55 as compared with 0.39 for single task tracking. A similar improvement was found for the CMT, with values of 0.55 and 0.46 respectively. Detailed analysis revealed that the correlation for tracking under dual-task conditions was suppressed by subjects with prior flying experience. They scored relatively low during the dual-task condition, but received good grades on the flight simulator. When these subjects ($n=8$) were omitted, the correlation increased from 0.55 to 0.74. Other scores were hardly affected by this correction. Prior flying experience probably decreases the validity of the simulator test by overemphasizing piloting skills, while dual-task performance in the laboratory does not seem to be biased by prior experience.

## Flight Training

Flight training for naval pilots, involves basic training, flying the Fokker Friendship and a conversion to the Lockheed Orion or the Westland Lynx helicopter. Obtaining an operational status as a first pilot was the relevant criterion for a successful completion. The performance for pilots who failed or were certified are depicted in Figure 7.

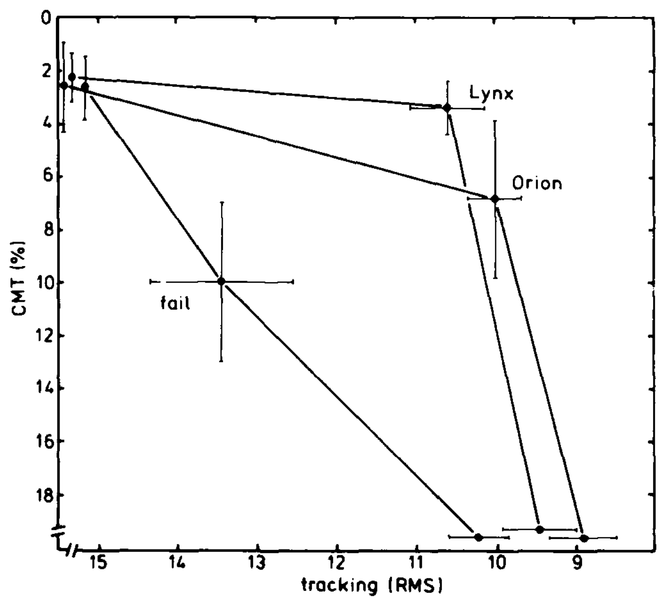



Figure 7. Single- and dual-task performance for pilots who passed or failed advanced training for Orion or Lynx. Standard errors indicated for both tasks.

Successful pilots are characterized by an excellent performance, particularly under dual-task conditions. It was only under dual-task conditions that reliable differences were found with the pilots who failed. Pilots who completed the training, performed better on tracking ($p<0.004$) as well as on the CMT ($p<0.037$). This finding confirmed that dual-task testing could provide a better discrimination of subjects who will have a chance of finishing training all the way. Failing pilots were characterized by a substantial performance decrement under dual-task conditions. It should be noted that all but one of the aspirants in this group had extensive flying experience when they entered basic selection. It seems that these subjects were allowed to enter the training program with the argument that prior experience would imply a better chance of success in advanced flight training. The present findings seem to indicate that such prior experience does not appear to be a critical factor in predicting success of operational training.

## Discussion

The present results suggest that pilot selection could be improved by testing aspirants under dual-task conditions. Under such conditions, performance was found to be related to the level of "pilot aptitude" as assessed by present selection, a flight simulator test and the results of advanced flight training. The tracking task was not influenced by prior flying experience and the CMT was successful in producing consistent task interference. Successful aspirants, now operational pilots, were characterized by excellent performance under both single- and dual-task conditions. Less successful aspirants performed less efficient under dual-task conditions as well as single task conditions, depending on how soon they failed in the selection and training process.

Trainability, i.e the ability to improve performance or automate information processing, was found to be important for the validity of the measures. Training resulted in a better discrimination between the groups of interest. Successful aspirants seem to be characterized by a fast rate of learning. This will serve to improve single- task as well as dual-task performance. Single task measures revealed differences or trends that were consistently confirmed by similar but more reliable differences in dual-task performance. Rapid learning and adjustment to the situation at hand is an important issue for pilot selection as it is known that military training is harsh and most often limited in the time allowed to reach a certain standard. Repeated task execution in a test procedure will, however, induce fatigue and could lead to "time on task" effects. The aspirants in this study were tested on a separate day and were instructed to be fit. These pilot candidates were also highly motivated and it was never observed that performance deteriorated as a function of training or fatigue. Preliminary results of similar experiments with university students revealed such a "time on task" effect. Dual-task studies are intended to push the human information processor "to the limits". An adequate motivation or incentive should always be considered as a critical factor for obtaining useful results during test development with other subjects.

It is evident from the present results, that prior flying experience cannot be used as a solid basis for selection, as it does not provide a guarantee for adequate dual-task performance. Pilot aptitude could be easily overestimated by the aviation psychologist or the instructor. Most of the pilots that didn't finish advanced training had a B3 license. It was only under dual-task conditions or operational flying that performance was found to be unsatisfactory. This observation confirms the original problem that pilots who were able to fly a Fokker Friendship, did not necessarily had the capacity or potential to obtain an operational status on a Lockheed Orion or Westland Lynx helicopter. Other studies also confirmed that pilots are not necessarily better under dual-task conditions (12). There are, however, several levels of pilot certification and experience to be considered. A study with the CMT and Dichotic Listening Test by Boer (13) showed that performance was better for Airline pilots as compared to pilots with a private certificate. Significant individual differences in dual-task performance are still to be expected for aspirants with a limited certification. A second problem of prior flying experience is that the validity of a simulator test as a selection tool will be limited. Dual-task testing could provide a tool that will contribute to minimize the risk of overestimating pilot aptitude for flying and operating under military conditions.

The experimental version of the dual-task test, was developed with a laboratory computer system. This prohibits the application of the tasks during standard selection. We therefore developed a PC-based version of the test procedure. This version has been named the "PILOT" test (Processing Information in Loading Or Time-sharing conditions). The applicability of this test for the selection of civil aviators is now under investigation. Both component tasks are integrated in a standardized battery of laboratory tasks known as the "Taskomat". This multi-purpose battery is based on well defined and analysed tasks from experimental psychology and was developed by Boer, Gaillard and Jorna (14).

**References**


1 Ackerman, P.L., Schneider, W. and Wickens, C.D., Deciding the existence of a time-sharing ability: A combined methodological and theoretical approach. Human Factors, 26, 71-82, 1984.

2 Braune, R. and Wickens, C.D., Time-sharing revisited: Test of a componential model for the assessment of individual differences. Ergonomics, 29, 1399-1414, 1986.

3 Norman, D.A. and Bobrow, D.J., On the analysis of performance operating characteristics. Psychological Review, 83, 508-510, 1975.

4 Shiffrin, R.M. and Schneider, W., Controlled and automatic human information processing II: Perceptual learning, automatic attending and a general theory. Psychological Review, 84, 127-190, 1977.

5 Navon, D. and Gopher, D., On the economy of the human processing system. Psychological Review, 86, 214-253, 1979.

6 Wickens, C.D., The structure of processing resources. In: Nickerson, R. and Pew, R. (Eds.), Attention and performance VIII, Erlbaum, New York, 1980.

7 Gopher, D., A selective attention test as predictor of success in flight training. Human Factors, 24, 173-183, 1982.

8 Damos, D.L. and Wickens, C.D., The identification and transfer of timesharing skills. Acta Psychologica, 46, 15-39, 1980.

9 Jorna, P.G.A.M., Stress, Information processing and diving. Report IZF 1981-4, TNO Institute for Perception, Soesterberg, The Netherlands, 1981.

10 Jorna, P.G.A.M. and Mulder, G., Mental load and the processing of information in the underwater environment. In: H. Ursin et al. (Eds.) Biological and psychological bases of psychosomatic disease. Pergamon Press, Oxford, 1983.

11 Hess, R.A., Pursuit tracking and higher levels of skill development in the human pilot. IEEE Transactions on Systems, Man and Cybernetics, 4, 262-273, 1981.

12 Tsang, P.S., Can pilots time-share better than non-pilots? Applied Ergonomics, 17.4, 284-290, 1986.

13 Boer, L.C., Attention tasks and their relation to aging and flight experience. Report IZF 1986-4, TNO Institute for Perception, Soesterberg, The Netherlands, 1986.

14 Boer, L.C., Gaillard, A.W.K. and Jorna, P.G.A.M., [Taskomat: A task battery for information processing] in Dutch. Report IZF 1987-2, TNO Institute for Perception, Soesterberg, The Netherlands, 1987.

# PREDICTING AIR COMBAT MANEUVERING (ACM) PERFORMANCE

by
G. R. Griffin
Naval Aerospace Medical Research Laboratory
NAS Pensacola
Pensacola Florida, 32508-5700
U.S.A.

## SUMMARY

A difficult aspect of predicting fleet pilot performance is acquiring meaningful and reliable, inflight criteria. This study was an attempt to predict Air Combat Maneuvering (ACM) performance using performance-based laboratory tests and to evaluate the VF-43 adversary squadron's grading of inflight ACM performance in the Fleet Fighter ACM Readiness Program at Naval Air Station (NAS) Oceana, Virginia Beach, VA.

In an initial evaluation (Study I), F-4 pilots performed in Fleet Fighter ACM Readiness exercises and completed performance-based perceptual motor and multitask tests. Results indicated that dichotic listening test measures, obtained during multitask conditions, could be used to reliably predict ACM inflight criteria. Results of a larger sample of F-14 pilots (Study II) indicated that an overall ACM grade (OAG) assigned by VF-43 adversary personnel can be predicted reliably by an objective kill difference composite score and three subjective measures: situational awareness, mutual support, and energy management. These four measures accounted for 78% of the variance with the OAG. A correlational analysis suggests that the VF-43 grading process is reliable and consistent.

## INTRODUCTION

Research is ongoing at the Naval Aerospace Medical Research Laboratory to predict fleet aviator inflight performance using perceptual psychomotor and information processing tasks. The goal is to develop relevant laboratory tasks, test aircrew, and relate aircrew test performance to simulated and actual flight performance. From this effort, it may be possible to aid decisions concerning aircrew selection, training pipeline assignment, and post-training aircraft assignment. Crucial to this research is the identification of useful, valid, and reliable measures of flight performance for the validity assessment of predictor tests.

Previous United States Navy research to predict operational performance has been encouraging (1-7). Rickus and Berkshire (4) reported that peer ratings obtained in Navy preflight training were useful in identifying successful and unsuccessful aviators in combat (Vietnam). Bale et al. (2) evaluated F-4 Replacement Air Group (RAG) training during the midsixties and developed a prediction equation that could reduce RAG attrition from 13.3 to 8.3%. A study of Tactical Aircrew Combat Training System (TACTS) F-4 air combat maneuvering by Ciavarelli et al. (1) in the late seventies found three measures (angle-off-tail, closing velocity, and indicated air speed) that were significantly related to ACM performance. Brictson et al. (3) were able to successfully predict F-4 carrier landing performance. Shannon et al. (7) found that a relatively small set of RAG measures can reliably predict final overall RAG grade (multiple $R$ = .84). The two most important measures (carrier qualification power/nose control and offensive ACM) accounted for 73% of the variance with the final overall RAG grade. In two subsequent studies, Shannon and Waag (6) found that an equation based on an east coast RAG reliably predicted performance of F-4 pilots on a west coast RAG and reported (5) that experience and seven undergraduate training grades reliably predicted final overall RAG grade (multiple $R$ = .51).

Despite these positive results, new aircraft and weapons system technology may have made previous research results obsolete. Consequently, the approaches of the previous studies and the present effort differ. Previous studies (2-6) used pencil-and-paper selection tests and undergraduate training measures to predict performance. This approach used performance-based tests of cognitive, perceptual, and multitask functioning to predict fleet operational aviator performance.

The present study represents an attempt to predict ACM inflight performance using performance-based automated tests and an evaluation of the VF-43 adversary squadron's grading of aircrew performance in the Fleet Fighter ACM Readiness Program at NAS Oceana. The purpose of the latter effort was to select useful and reliable criteria for ACM performance assessment and validation of future laboratory tests. In addition, correlations between measures of the TACTS and vision tests were determined.

### Study I. Multitask Test and Marine Pilot ACM Performance

The purpose of this evaluation was to test the feasibility of predicting ACM performance with performance-based perceptual-motor and cognitive multitask tests.

SUBJECTS

Twenty-two F-4 pilots from Marine Squadron 451 served as subjects during their participation in a Fleet Fighter ACM Readiness Program exercise at NAS Oceana during the summer of 1985.

PROCEDURE

Air combat maneuvering performance data are routinely collected at the NAS Oceana TACTS facility. The data are used by VF-43 adversary squadron personnel to develop aircrew and squadron ACM performance ratings. The performance ratings provide training feedback to individual aircrews by highlighting their strengths and weaknesses in ACM and also provide a method for military managers to assess overall squadron readiness. In addition, the TACTS Fleet Fighter ACM Readiness Exercise results serve as a base to evaluate the tactical employment of aircraft and weapon systems. Typical data resulting from the readiness exercises are presented in Table 1. A description of the TACTS training system and definitions for specific TACTS performance measures are in appendixes A and B, respectively.

Eighteen of the 22 Marine pilots completed single- and multitask cognitive and perceptual-motor tests during the readiness exercise. The tests consisted of a 24-trial dichotic listening task (DLT) followed by 6-min performance on a psychomotor task. Both tasks were then performed simultaneously. Correlational and multiple regression analyses were conducted on the ACM performance measures to identify suitable criteria and to evaluate the strengths of the correlations among the various measures. Subjects' test performance data were correlated with the identified criteria.

RESULTS

Pearson correlations of 27 measures associated with the VF-43 adversary squadron's evaluation of ACM performance of 22 F-4 pilots are presented in Table 1. The overall ACM grade (OAG) was significantly related to offensive maneuvering ($r = .67$), situational awareness ($r = .81$), and mutual support ($r = .56$). In addition, the OAG was significantly related to mean time to first kill ($r = -.42$), number of VF-43 adversary squadron missile shots ($r = -.65$), and the number of times a pilot was "killed" ($r = -.70$) in the simulated exercises. The negative correlations indicated that a higher ACM grade was associated with shorter times to first kill, fewer adversary squadron missile shots taken, and fewer times being "killed" in the simulated exercises (better ACM performance).

A multiple regression analysis indicated that situational awareness, offensive maneuvering, number of times killed, and mutual support accounted for 89% of the variance associated with the OAG criterion ($R = .95$, $F(4, 17) = 36.19$, $p < .0001$). The OAG and these four measures were then correlated with the single- and multitask cognitive and psychomotor test performance of 18 of the 22 pilots.[1] A derived composite kill-difference score (the total number of ACM kills minus the number of times a pilot was killed in the TACTS simulated exercises) was included in the correlational analyses as well as the total number of flight hours, which ranged from 337 to 1925. Of the 42 correlations computed between the tests and ACM criteria, 4 (10%) were significant at the .05 or .01 level of confidence.

The Pearson correlations of the cognitive and psychomotor tests shown in Table 2 indicated that a DLT measure (DLT-1) obtained during multitask performance was significantly related to offensive maneuvering ($r = .62$) and the kill-difference composite score ($r = .49$). A DLT multitask measure based on a slightly different scoring procedure (DLT-2) was significantly related to the OAG ($r = .49$) and the offensive maneuvering score ($r = .60$). Number of flight hours was unrelated to any ACM or test performance measure. These results, although based on a small sample of Marine pilots, support the feasibility of developing a series of performance-based cognitive, perceptual, and multitask tests to predict aviator performance.

A series of important questions concerning the Fleet Fighter ACM Readiness Program evaluation process was unresolved: Are the resulting grades reliable for Navy pilots flying contemporary F-14 aircraft? What is the relation of the VF-43 grading process to more objective TACTS ACM performance measures (i.e., total number of kills, visual identification (VID) range, VID kills, and engaged kills)? Which ACM measures are most predictive of ACM performance?

**Study II: Fleet Fighter ACM Readiness Program Grades as Criteria**

The purpose of the second study was to answer the questions posed above and assess the utility of VF-43 ACM grades and TACTS ACM performance measures as criteria for the validation of tests developed to predict ACM performance.

SUBJECTS

Subjects were 125 Navy F-14 pilots (10 fighter squadrons) who participated in the Fleet Fighter ACM Readiness Program against the VF-43 adversary squadron at NAS Oceana during 1985 and January of 1986.

---

[1] Four pilots did not volunteer to complete the series of tests.

TABLE 1. Pearson Correlations of Individual ACM Performance Measures[a] with Overall ACM Grade[b] ($N$ = 22).

| Subjective measures | $r$ |
|---|---|
| Use of environment | .05 |
| Start/VID start | .07 |
| First move | -.18 |
| Aggressiveness | .16 |
| Offensive maneuvering | .67** |
| Defensive maneuvering | -.04 |
| Keeping sight/lookout | .39 |
| Energy management | .23 |
| Mental plot | -.17 |
| Situational awareness | .81** |
| Bugout technique | -.13 |
| Weapon system employment | .24 |
| VID technique | -.08 |
| VID communication | .05 |
| UHF communication | .24 |
| Game plan usage | .16 |
| Mutual support | .56** |
| Debrief | -.17 |
| Reconstruction | -.37 |
| **Objective measures** | **$r$** |
| Total number TACTS kills | .31 |
| Number of missiles launched | .16 |
| Mean time (s) to first kill | -.42* |
| Visual tally-ho mean range | .01 |
| VID mean range | -.05 |
| Number of times killed | -.70** |
| Number VF-43 missile shots | -.65** |

[a]Performance measure definitions are in appendix B.

[b]The overall grade is a composite of the 19 subjective measures.

* $p < .05$
** $p < .01$

TABLE 2. Correlation of ACM Performance Criteria and Single and Multitask Cognitive and Psychomotor Tests ($n$ = 18 F-4 Pilots).

| Measures | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Overall ACM grade | | | | | | | | | | | | |
| 2. Situational awareness | .83** | | | | | | | | | | | |
| 3. Offensive maneuvering | .83** | .61** | | | | | | | | | | |
| 4. Number times killed | -.75** | -.58* | -.58* | | | | | | | | | |
| 5. Mutual support | .51* | .58* | .25 | -.21 | | | | | | | | |
| 6. Kill-difference score | .50* | .25 | .41 | -.59** | -.03 | | | | | | | |
| 7. Flight hours | .13 | -.01 | .06 | -.39 | .21 | .05 | | | | | | |
| 8. Multitask DLT 1 | .43 | .10 | .62** | -.27 | -.16 | .49* | .11 | | | | | |
| 9. Multitask DLT 2 | .49* | .12 | .60** | -.33 | -.03 | .37 | .27 | .93** | | | | |
| 10. Multitask PMT | -.14 | .04 | -.25 | .13 | .20 | -.05 | -.18 | -.40 | -.45 | | | |
| 11. Single task DLT 1 | .21 | -.02 | .28 | -.17 | -.15 | .35 | .04 | .56* | .53* | -.72** | | |
| 12. Single task DLT 2 | .32 | .15 | .36 | -.20 | -.19 | .22 | -.10 | .60** | .55** | -.67** | .85** | |
| 13. Single task PMT | .10 | -.03 | .16 | -.18 | -.29 | .16 | .07 | .27 | .29 | .31 | -.12 | -.06 |

*.05 = .47
**.01 = .59

## PROCEDURE

Air combat maneuvering "competitive exercise" performance data were collected for the Navy participants in the Fleet Fighter ACM Readiness Program and analyzed to derive correlational statistics. Multiple regression analyses were performed to study the relative importance of specific predictors and to derive criteria that would predict the overall ACM grade.

## RESULTS

Performance measure descriptive statistics and Pearson correlations between the OAG and 19 subjective and 12 objective TACTS measures are presented in Table 3. The kill-difference composite score was the measure most highly correlated with OAG ($r$ = .76), followed by the engaged kill-difference composite score ($r$ = .70). The total number-of-kills measure was related to OAG ($r$ = .65), as were missiles launched ($r$ = .58), VID-kill ($r$ = .39), and engaged-kill ($r$ = .57) scores. As expected, the number-of-times-killed measure was significantly and negatively related to OAG ($r$ = - .51). Number of radar locks was significantly correlated with the OAG ($r$ = .24) as well. Four of the objective measures--mean time-to-first-kill, radar lock mean range, visual tally-ho mean range, and VID mean range measures--were not significantly related to OAG.

Surprisingly, the mean time-to-first-kill score was unrelated to OAG. One explanation for this result may be that the time-to-first-kill score is an average of both VID and engaged-kill times. This pooling of relatively short (VID) and longer (engaged) kill times may have a confounding effect on the resulting correlations. Separation of VID and engaged kill times might enable a better understanding of the relation of this ACM score with the OAG and other TACTS measures.

An examination of the subjective measures, as shown in Table 3, indicated that 11 measures were significantly correlated with the OAG. Situational awareness (described by VF-43 adversary personnel as a synonym for ACM proficiency) correlated most highly with OAG ($r$ = .70), followed by offensive maneuvering ($r$ = .53), aggressiveness ($r$ = .45), mutual support ($r$ = .44), and start/VID start ($r$ = .40). Defensive maneuvering, keeping sight, energy management, weapon system employment, VID technique, and game plan measures were significantly correlated with the OAG as well, with correlations between .23 and .39. Those measures not significantly correlated with the OAG were use of environment, first move, mental plot, bugout technique, VID communication, UHF communication, debrief, and reconstruction.

## REGRESSION ANALYSIS

To examine which subjective and objective measures would best predict the OAG, a series of forward selection multiple regression analyses (8) was conducted. A forward selection stepwise multiple regression technique was used because of multicollinearity (high intercorrelations) among certain of the objective and subjective measures. The first regression (appendix C, Table C-1) was based on the subjective measures in Table 3. Results indicated that a six-measure regression model accounted for 83% of the variance with OAG ($R$ = .91, $F(6, 118)$ = 98.57, $p$ < .0001). The situational awareness measure entered the regression first and accounted for 49% of the variance with the OAG. Offensive maneuvering, mutual support, start/VID start, energy management, and keeping sight measures then entered the regression equation accounting for 11, 7, 8, 4, and 4% additional variance, respectively.

A second regression analysis (appendix C, Table C-2) was conducted using the objective performance measures in Table 3. The measures total TACTS kills, kill-difference score, and engaged kill-difference score were excluded because they represented combinations of other measures. Number of missiles launched was related to total TACTS kills ($r$ = .83), engaged-kills ($r$ = .67), VID-kills ($r$ = .57), and the kill-difference score ($r$ = .78). Since this measure is simply a means of achieving TACTS kills, it too was excluded. These composite and/or duplicative measures were omitted from the regression to increase insight into those specific performance measures most important to the OAG. The results of the multiple regression indicated that engaged-kills, number-of-times-killed, and VID-kills accounted for 62% of the variance with the OAG. The engaged-kill measure entered the model first and accounted for 33% of the variance associated with the OAG. Number-of-times-killed and VID-kill measures followed in succession, accounting for 19 and 10% additional variance, respectively. Finally, the mean time-to-first-kill measure entered the regression model but accounted for only 1% additional variance ($R$ = .79, $F(4, 120)$ = 51.02, $p$ < .0001). Both engaged-kills and VID-kills entered the regression model (both are significantly correlated with OAG, but the correlation between the two measures is low, $r$ = .14). These results suggest that the VID-kill and engaged-kill performance measures are statistically independent in this population of Navy pilots. They also emphasize the importance of pilot training in both of these ACM skills.

A third multiple regression model (appendix C, Table C-3) was based on a kill-difference score (a composite of the first three measures entering the second regression model) and the subjective measures of Table 3. The kill-difference measure entered the regression first, accounting for 57% of the variance with OAG. Next, the situational awareness measure entered the regression, accounting for an additional 14% of variance, followed by energy management and mutual support, which each contributed about 4% additional variance ($R$ = .89, $F(4, 120)$ = 109.39, $p$ < .0001).

These results, indicating the importance of kills in the VF-43 adversary squadron's grading of ACM performance, were expected since kill ratios from the competitive exercises of the Fleet Fighter Readiness Program represent a basic component of the grading process (9). Unexpectedly, situational awareness and other subjective measures contribute important additional variance in the prediction of OAG. Apparently, human judgment of ACM proficiency is an important element in these performance evaluations.

TABLE 3. Performance Measure Descriptive Statistics and Pearson Correlations Between Overall ACM Grade and ACM Performance Measures ($N$ = 125).

| Subjective measures | $r$ | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Overall grade | -- | 2.01 | .05 | 1.90 | 2.15 |
| Use of environment | .00 | 2.01 | .05 | 1.83 | 2.20 |
| Start/VID start | .40** | 2.04 | .16 | 1.67 | 2.88 |
| First move ($n$ = 113) | .12 | 2.01 | .18 | 1.50 | 2.50 |
| Aggressiveness | .45** | 2.08 | .13 | 1.88 | 2.75 |
| Offensive maneuvering | .53** | 2.08 | .18 | 1.67 | 2.50 |
| Defensive maneuvering | .39** | 1.96 | .14 | 1.56 | 2.25 |
| Keeping sight/lookout | .35** | 1.95 | .14 | 1.50 | 2.29 |
| Energy management | .37** | 2.00 | .15 | 1.57 | 2.38 |
| Mental plot | -.11 | 1.98 | .11 | 1.50 | 2.90 |
| Situational awareness | .70** | 1.93 | .25 | 1.25 | 2.75 |
| Bugout technique | .11 | 2.03 | .17 | 1.67 | 2.50 |
| Weapon system employment | .30** | 2.06 | .19 | 1.25 | 2.50 |
| VID technique | .30** | 1.99 | .12 | 1.67 | 2.33 |
| VID communication | -.08 | 1.99 | .05 | 1.75 | 2.17 |
| UHF communication | .12 | 2.00 | .08 | 1.50 | 2.25 |
| Game plan usage | .32** | 2.08 | .18 | 1.50 | 2.50 |
| Mutual support | .44** | 1.95 | .26 | 1.33 | 2.50 |
| Debrief ($n$ = 112) | .11 | 2.01 | .07 | 2.00 | 2.50 |
| Reconstruction | .16 | 2.01 | .04 | 2.00 | 2.25 |
| **Objective measures** | $r$ | **Mean** | **SD** | **Min** | **Max** |
| Total number TACTS kills | .65** | 5.63 | 2.80 | 0.0 | 14.0 |
| Number of missiles launched | .58** | 13.14 | 6.85 | 1.0 | 31.0 |
| Number of VID kills | .39** | 2.66 | 1.64 | 0.0 | 7.0 |
| Number of engaged kills | .57** | 2.97 | 2.05 | 0.0 | 10.0 |
| Mean time (s) to first kill (n =124) | -.02 | 42.98 | 25.31 | 6.0 | 115.3 |
| Number of radar locks | .24** | 4.95 | 1.55 | 0.0 | 7.0 |
| Radar locks mean range | -.04 | 13.82 | 3.22 | 0.0 | 26.0 |
| Visual tally-ho mean range | .12 | 2.80 | 1.37 | 0.0 | 6.5 |
| VID mean range | .16 | 1.59 | 0.88 | 0.0 | 5.2 |
| Number of times killed | -.51** | 1.68 | 1.10 | 0.0 | 5.0 |
| Kill-difference score | .76** | 3.92 | 3.14 | -2.0 | 13.0 |
| Engaged-kill-difference score | .70** | 1.29 | 2.44 | -3.0 | 9.0 |

** $p < .01$

## SITUATIONAL AWARENESS

At NAS Oceana, VF-43 adversary personnel define situational awareness as the "total of ACM." This definition seems appropriate based on the results reported here. Table 3 shows that situational awareness is the subjective measure most strongly related to OAG ($r$ = .70).

Those measures most strongly related to situational awareness, in addition to the OAG, are the kill-difference score, engaged kill-difference score, number-of-times-killed, VID-kills, engaged-kills, total TACTS kills, and number of missiles launched. Subjective measures--start/VID start, aggressiveness, offensive maneuvering, defensive maneuvering, keeping sight. VID technique, game plan and mutual support--are also significantly related to the situational awareness measure.

Objective measures unrelated to situational awareness (in this analysis) are visual tally and VID range, number of radar locks, radar lock range, and mean time-to-first-kill. Subjective measures that are not significantly related to situational awareness are environment, first move, energy management, mental plot, bugout, weapon employment, VID and UHF communication, debrief, and reconstruction.

## PERFORMANCE MEASURE RELIABILITY

An important aspect of this study concerns the reliability or consistency of the VF-43 performance measures. To evaluate the reliability of the TACTS objective performance scores and the more subjective VF-43 scoring process, the Navy pilot sample was randomly divided in half and performance measures were correlated with the OAG (Table 4). Subjects were divided on the basis of even/uneven chronological subject number.[2] Table 4 includes Pearson correlations based on the total sample to allow comparison with the correlations of both subsamples. In addition, the absolute difference of the Pearson correlations is presented. Table 4 reveals remarkably similar results, especially for the more objective TACTS parameters. The one objective measure that indicated a major correlation change was the mean time-to-first-kill measure, with an $r$ of .09 for the even and -.12 for the uneven subsample, an absolute difference of .21. This change in correlational value is not significant at the .05 level. Moreover, the mean time-to-first-kill measure is not significantly related to the OAG. All other objective measure correlational values were highly similar. Six of the subjective measures had an absolute difference correlation value of .20 or greater.

Only two of these measures, UHF communication and reconstruction, represented a significant difference between the even and uneven pilot subsamples ($p < .05$), based on a Fishers Z test of significant differences of correlations. Neither of these correlational values, however, was significantly correlated with OAG for the total sample or the two subsamples. In summary, the objective and subjective measures most highly correlated with OAG differ only slightly for the two randomly derived samples.

A second approach to establishing the reliability of the OAG was to apply the regression model of appendix C, Table C-3 (based on the kill-difference, situational awareness, energy management, and mutual support measures) to various pilot subsamples. This particular regression model was used because it represents the best prediction of OAG using both objective and subjective performance measures. A Pearson correlation value was computed between the predicted and actual OAG grade for eight different pilot subsamples (Table 5). Based on a Fishers Z transformation, the average of the eight correlation values is .88.

In summary, regardless of the sampling procedure, the model for predicting the OAG provided similar results. Because the grading of the aircrews by different adversary pilots seems consistent, we can assume that the internal criteria by which the grades are assigned are similar across adversary pilots. In essence, the grading process appears reliable.

PILOTS, AIRCREW, AND VISUAL PERFORMANCE

Although the F-14 aircraft normally employs both a pilot and a Radar Intercept Officer (RIO) working as a team, this study addressed those measures associated with pilot performance. The VF-43 scoring process emphasizes pilot proficiency, since the pilot maneuvers the aircraft and fires the missiles, and, as the aircraft commander, is responsible for engagement outcome. However, the RIO's efforts in operating the radar, keeping sight, and performing lookout also have an important effect on ACM engagement outcome. Consequently, two RIO measures (number of radar locks and radar lock mean range) were included in this analysis to examine the relation of RIO performance to pilot tally-ho and pilot aircraft identification range (important in pilot visual performance). The importance of the RIO's radar skills and pilot visual performance are demonstrated by the significant Pearson correlations between the number of radar locks, visual tally-ho mean range, VID mean range, and other objective TACTS ACM scores of Table 6.

Initially, it was unclear as to why an RIO performance parameter (number of radar locks) would be significantly related to pilot visual tally-ho ($r = .43$) and VID mean range ($r = .50$). One possibility is that a radar lock acts to decrease the visual target search area for the pilot, who then can attend to the diamond,[3] knowing that an adversary aircraft will ultimately become a visual target at the indicated location on the head-up display.

A radar lock is important to kills, and it is a requirement for successful use of a forward aspect missile. Number of radar locks was significantly correlated with VID kills ($r = .42$) and, to a lesser extent, engaged kills ($r = .22$). The number of radar locks was significantly related to total number of TACTS kills ($r = .41$), number of missiles launched ($r = .40$), and the kill-difference score ($r = .34$). Radar lock mean range was not a significant predictor for the majority of objective ACM performance measures. Apparently, when radar lock is accomplished, it occurs at distances so great that the variability in lock ranges does not influence subsequent ACM performance. Failure to acquire radar lock is another matter, however, as noted above.

Visual tally-ho mean range and VID mean range, as previously noted, are strongly correlated with the radar lock measure. Apparently, a radar lock significantly enhances the pilot's acquisition of visual targets. Since VID of adversary aircraft is required before missile launch, under present tactical rules, vision-dependent ACM performance measures should be positively related to number of TACTS kills. Our data support this hypothesis. That is, a greater visual tally-ho range and greater VID range are each significantly associated with a greater number of TACTS kills ($r = .30$ and .41, respectively). Further, it was hypothesized that vision-dependent ACM performance might be more highly related to the number of VID kills rather than engaged kills.[4] Our data support this hypothesis. Visual tally-ho performance is significantly correlated with VID kills ($r = .45$) but not engaged kills ($r = .05$). Visual identification performance is also significantly related to VID kills ($r = .48$) and not engaged kills ($r = .16$).

---

[2] Subject performance data were ordered for statistical analysis by date of the ACM readiness evaluation and the alphabetical order of pilot name.

[3] An area of the head-up display, delineated as a diamond shape, indicating the location of the radar target.

[4] VID kills are those that occur immediately following initial target detection and identification and are generally made with radar directed missiles fired head-on. Engaged kills occur during subsequent dogfighting, when pilots attempt to maneuver behind their adversary to fire guns or heat seeking missiles.

TABLE 4. Pearson Correlations for Total, Even, and Uneven Ordered Pilots, and Correlation Absolute Difference Scores.

| Performance measure correlation with OAG | All pilots (N = 125) | Even (n = 62) | Uneven (n = 63) | Difference |
|---|---|---|---|---|
| Use of environment | .00 | .03 | -.03 | .06 |
| Start/VID start | .40** | .38** | .42** | .04 |
| First move | .12 | .05 (n=56) | .20 (n=57) | .15 |
| Aggressiveness | .45** | .41** | .47** | .06 |
| Offensive maneuvering | .53** | .49** | .57** | .08 |
| Defensive maneuvering | .39** | .38** | .41** | .03 |
| Keeping sight/lookout | .35** | .45** | .26* | .19 |
| Energy management | .37** | .46** | .28* | .18 |
| Mental plot | -.11 | -.13 | -.06 | .07 |
| Situational awareness | .70** | .63** | .76** | .13 |
| Bugout technique | .11 | .22 | .01 | .21 |
| Weapon system employment | .30** | .39** | .18 | .21 |
| VID technique | .30** | .29* | .34** | .05 |
| VID communication | -.08 | .02 | -.16 | .18 |
| UHF communication | .12 | -.15 | .33** | .48**(1) |
| Game plan usage | .32** | .42** | .22 | .20 |
| Mutual support | .44** | .38** | .49** | .11 |
| Debrief | .11 | .10 (n=55) | .12 (n=57) | .02 |
| Reconstruction | .16 | .32** | -.10 | .42*(1) |
| Total number TACTS kills | .65** | .68** | .61** | .07 |
| Number missiles launched | .58** | .60** | .56** | .05 |
| Number of VID kills | .39** | .33** | .45** | .12 |
| Number of engaged kills | .57** | .63** | .51** | .12 |
| Mean time-to-first-kill | -.02 | .09 (n=61) | -.12 (n-63) | .21 |
| Number of radar locks | .24** | .23 | .26* | .03 |
| Radar locks mean range | -.04 | .02 | -.08 | .10 |
| Visual tally-ho mean range | .12 | .04 | .17 | .13 |
| VID mean range | .16 | .07 | .22 | .15 |
| Number of times killed | -.51** | -.47** | -.56** | .09 |
| Kill-difference score | .76** | .79** | .73** | .06 |
| Engaged kill-difference score | .70** | .74** | .67** | .07 |

* $p < .05$
** $p < .01$
(1) Fisher Z test of significance of Pearson correlations

TABLE 5. Pearson Correlation Values for Eight Pilot Subsamples Based on Predicted and Actual OAG.

| Pilot subsamples | r* | Number |
|---|---|---|
| Even | .89 | 62 |
| Uneven | .88 | 63 |
| First half | .84 | 62 |
| Second half | .90 | 63 |
| 1st, 3rd quarter | .88 | 62 |
| 2nd, 4th quarter | .89 | 63 |
| 2nd, 3rd quarter | .89 | 62 |
| 1st, 4th quarter | .89 | 63 |

*All values significant, $p < .01$

TABLE 6. Pearson Correlations Between Radar Locks, Pilot Visual Tally and VID, and Objective ACM Performance (N = 125 Navy Pilots).

| ACM objective measure | Number of radar locks | Radar lock mean range | Visual tally-ho mean range | VID mean range |
|---|---|---|---|---|
| Total number TACTS kills | .41** | -.10 | .30** | .41** |
| Number missiles launched | .40** | -.21* | .32** | .42** |
| Number of VID kills | .42** | .12 | .45** | .48** |
| Number of engaged kills | .22* | -.05 | .05 | .16 |
| Mean time-to-first-kill | -.12 | -.01 | -.19* | -.14 |
| Number of radar locks | - | .25** | .43** | .49** |
| Radar lock mean range | .25** | - | .05 | -.07 |
| Visual tally-ho mean range | .43** | .05 | - | |
| VID mean range | .50** | -.07 | .68** | - |
| Number of times killed | .06 | .00 | .10 | .07 |
| Kill-difference score | .34** | -.09 | .23** | .33** |
| Engaged kill-difference score | .14 | -.05 | .00 | .11 |

* $p < .05$
** $p < .01$

Number of radar locks, visual tally-ho mean range, and VID mean range are negatively related to the mean time-to-first-kill measure. These correlations are negative since the launch of a forward aspect missile (the best means of achieving a quick kill) generally depends on achieving each of these measures in sequence. Having a longer tally-ho or VID range enables better aircrew preparation at the merge and reduces time-to-first-kill. Additionally, a radar lock allows more certainty in visual search and produces longer range visual target acquisition and aircraft identification. In summary, the relation of radar locks to improved vision-dependent ACM performance and the relation of visual tally-ho and VID performance to subsequent missile launch and a VID kill reflect a necessary sequence of performance events for achieving mission success.

EXPERIENCE AND ACM PERFORMANCE

Table 7 presents correlations between ACM performance criteria and measures of experience--specifically, age, jet hours, total jet hours, TACTS hours, and total ACM flight hours. Of 45 correlations, 8 (18%) were significant at the .05 or .01 level of confidence. The one criterion consistently related to age or flight experience measures was the mean time-to-first-kill score, which produced significant correlations with age ($r$ = -.34), jet hours ($r$ = -.36), total flight hours ($r$ = -.32), and TACTS hours ($r$ = -.22). In each case, greater age or more flight experience was associated with shorter mean time-to-first-kill scores (better performance). Visual ID performance was significantly related to jet hours ($r$ = .24) and total ACM flight hours ($r$ = .21). Only one experience measure, total ACM flight hours, was significantly related to the OAG ($r$ = .23), VID kills ($r$ = .21), and the kill difference score ($r$ = .21). These correlations were small, accounting for variance of only 4 or 5%. There were no significant correlations between age or flight experience measures and situational awareness, visual tally range, VID range, number of times killed, and the number of engaged kills.

TABLE 7. Pearson Correlations for Experience Measures and ACM Performance ($n$ varies).

| ACM performance criteria | Age $n$=89 | Jet hours $n$=88 | Total hours $n$=88 | TACTS hours $n$=85 | Total ACM $n$=85 |
|---|---|---|---|---|---|
| OAG | .10 | .20 | .17 | .13 | .23* |
| Situational awareness | .04 | .08 | .06 | -.02 | .08 |
| Mean time-to-first-kill | -.34** | -.36** | -.32** | -.22* | -.16 |
| Visual tally mean range | .05 | -.01 | -.04 | .06 | -.01 |
| Visual ID mean range | -.02 | -.06 | -.11 | -.08 | .06 |
| Number of times killed | -.06 | -.05 | .03 | .14 | -.02 |
| VID kills | .09 | .24* | .17 | .16 | .21* |
| Engaged kills | .13 | .10 | .13 | -.02 | .13 |
| Kill-difference score | .06 | .10 | .07 | .04 | .21* |

* $p < .05$
** $p < .01$

Our results indicate that ACM experience influences ACM performance, especially in achieving VID kills and improved time-to-first-kill scores. On the other hand, these results also suggest that experience is <u>not</u> related to situational awareness, visual tally, VID range, engaged kill, or being killed criteria. Future evaluations of ACM TACTS performance should include experience factors similar to those examined here to better understand the relation of age and experience to TACTS ACM performance.

## DISCUSSION AND CONCLUSIONS

### Study I. Multitask Test and Marine Pilot ACM Performance.

This evaluation was conducted to test the feasibility of predicting ACM performance with perceptual motor and cognitive multitask tests. Eighteen F-4 pilots performed in Fleet Fighter ACM Readiness exercises and completed automated performance-based tests.

Initial analyses indicated that the overall ACM grade (OAG) associated with the VF-43 adversary squadron's evaluation of ACM performance of F-4 pilots was significantly and positively related to offensive maneuvering, situational awareness, and mutual support measures. In addition, the OAG was significantly but negatively related to the objective TACTS measures, mean time-to-first-kill, adversary squadron missile shots, and the number of times a pilot was "killed" in the simulated exercises. The negative correlations indicated that a higher ACM grade was associated with shorter times-to-first kill, fewer adversary squadron missile shots taken, and fewer times being "killed" in the simulated exercises. A multiple regression analysis indicated that four of these measures, situational awareness, offensive maneuvering, number of times killed, and mutual support, could reliably predict the OAG criterion.

The OAG and the four criterion measures were then correlated with the single- and multitask cognitive and psychomotor test performance of the F-4 pilots. A derived composite kill-difference score, based on the total number of ACM kills less the number of times a pilot was killed in the TACTS simulated exercises, was included in the correlational analyses as well. A DLT measure obtained during multitask performance was significantly related to offensive maneuvering and the kill-difference composite score. A DLT multitask test measure based on a slightly different scoring procedure was significantly related to the OAG and the offensive maneuvering score.

Conclusion: Multitask tests reliably predicted ACM performance for a small sample ($n$ = 18) of F-4 pilots.

Unresolved, however, was a series of important questions concerning the Fleet Fighter ACM Readiness Program evaluation process: Are the resulting grades reliable for Navy pilots flying contemporary F-14 aircraft? What is the relation of the VF-43 grading process to more objective TACTS ACM performance measures (i.e., total number of kills, VID range, VID kills, and engaged kills)? Which ACM measures are most predictive of ACM performance?

**Study II: Fleet Fighter ACM Readiness Program Grades as Criteria.**

Objectives of Study II were to identify criteria for the validation of tests designed to predict ACM performance and estimate the reliability of readiness grades used to assess Navy pilot ACM proficiency.

An examination of subjective and objective measures of the TACTS ACM competitive exercise performance of 125 naval aviators participating in Fleet Fighter ACM Readiness Program Exercises at NAS Oceana, indicated that the OAG can be reliably predicted by a relatively few measures. These were a kill-difference composite score, situational awareness, energy management, and mutual support measures.

A separate correlational analysis examined the reliability of the Fleet Fighter ACM grading process. Subjects were randomly divided into two groups and the correlations between the various performance measures and OAG were examined. The resulting $r$ values were highly similar. In addition, correlation values were computed for eight different pilot subsamples based on the predicted OAG and the actual OAG. The average of the eight correlation values was .88. These results suggest that the Fleet Fighter Readiness grading process is reliable. Regardless of the subject sampling procedure, the model for predicting OAG provided highly similar results. Apparently, the grading of ACM performance by different VF-43 adversary pilots was consistent.

Conclusion: Fleet Fighter ACM Readiness program grades are reliable and suitable criteria for validating tests designed to predict F-14 pilot ACM performance.

Additional correlations were computed between ACM performance criteria and measures of experience--specifically, age, jet hours, total jet hours, TACTS hours and total ACM flight hours. The one criterion consistently related to age or flight experience measures was the mean time-to-first-kill score, which produced significant correlations with age, jet hours, total flight hours, and TACTS hours. In each case, greater age or more flight experience was associated with shorter mean time-to-first-kill scores (better performance). Visual ID kill performance was significantly related to jet hours and total ACM flight hours. Total ACM flight hours was significantly related to the OAG, VID kills, and the kill-difference score. There were no significant correlations between age or flight experience measures and situational awareness, visual tally range, VID range, number of times killed and the number of engaged kills.

Conclusion: Experience in ACM influences performance, especially in achieving VID kills and improved time-to-first-kill scores. Experience in ACM was <u>not</u> related to situational awareness, visual tally, VID range, engaged kill, or being-killed criteria.

## RECOMMENDATIONS

Results (Study I) demonstrated the feasibility of using automated, synthetic, cognitive, perceptual, and multitask tests to predict TACTS F-4 pilot ACM proficiency and indicated (Study II) that Fleet Fighter ACM Readiness Program grades are reliable criteria for validating tests designed to predict ACM performance.

To achieve the goal of validating tests to aid in aircrew selection and assignment decisions, the following research is recommended:

1. Synthetic cognitive, perceptual, and multitask tests should be administered to a suitable sample of F-14 pilots performing in Fleet Fighter Readiness Evaluations to replicate initial test results.

2. Pilot experience data should be included in the above effort to study the relation of age and experience to TACTS ACM performance.

The successful validation of synthetic tests to predict ACM performance would be valuable for improving the quality and capabilities of fighter aircrew through their initial selection and subsequent assignment to training pipelines and aircraft.

## REFERENCES

1. Ciavarelli, A, Brictson, C. A., and Young, P. A., Development and Application of Performance Criteria and Aircrew Assessment Methods for the Air Combat Maneuvering Range (ACMR), NAMRL Special Report 79-5, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1979.

2. Bale, R. M., Ric'us, G. M., and Ambler, R. K., Replacement Air Group Performance as a Criterion for Aviation Training, NAMRL-1126, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1973.

3. Brictson, C. A., Burger, W. J. and Gallagher, T. J., "Prediction of Pilot Performance during Initial Carrier Landing Qualifications." Aerospace Medicine, Vol. 43, pp. 483-487, 1972.

4. Rickus, G. M. and Berkshire, J. R., Development of an Aviation Combat Criterion, NAMI-1047, Naval Aerospace Medical Institute, Pensacola, FL, 1968.

5. Shannon, R. H. and Waag, W. L., The Prediction of Pilot Performance in the F-4 Aircraft, NAMRL-1186, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1973.

6. Shannon, R. H. and Waag, W. L., Toward the Development of a Criterion for Fleet Effectiveness in the F-4 Fighter Community, NAMRL-1173, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1973.

7. Shannon, R. H., Waag, W. L., and Ferguson, J. C A New Approach to Criterion Development in the Replacement ... Group, NAMRL-1158, Naval Aerospace Medical Research Laborato..., Pensacola, FL, 1973.

8. SAS User's Guide: Statistics (5th ed.), SAS Institute, Inc. Cary, NC, 1985.

9. Commander Fighter Wing One Instruction 3710.3E, Naval Air Station Oceana, Virginia Beach, VA, 11 Feb 1985.

## ACKNOWLEDGMENTS

The research reported in this paper was completed under the Naval Medical Research and Development Command work unit 63706N M0096001.0151. The views expressed in this article are those of the author and do not reflect the official policy or position of the Navy, Department of Defense, nor the U. S. Government.

Volunteer subjects were recruited, evaluated, and employed in accordance with the procedures specified in Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based on voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines.

## APPENDIX A

The Tactical Aircrew Combat Training System (TACTS) is one of the most technologically sophisticated training systems in existence. The system is a computer based tracking and data communication network that enables air combat maneuvering (ACM) training and simulated weapons firing of actual aircraft engaged in ACM in real time. The TACTS system provides military managers with visual flight dynamics, weapons system status, and weapons firing information. All data (including the visual representation of aircraft) are recorded on magnetic tape for use in the debriefing of both adversary and fighter aircrews. The need for such a training system became apparent during the Vietnam conflict in which pilots often failed to recognize when they were in a correct firing envelope for the missile weaponry of that day. The TACTS system enables the employment of both rear-aspect and forward-aspect missile weaponry and serves as a means of evaluating the tactical use of both missile and aircraft weapon systems in simulated air combat.

## APPENDIX B

### ACM PERFORMANCE MEASURE DEFINITIONS

1. OVERALL ACM GRADE - a composite of 19 subjective measures (see Table 1).

2. ENVIRONMENT - use of weather conditions to gain an advantage in ACM.
3. START/VID START - position at start of engagement when the fighter and adversary aircraft merge.*
4. FIRST MOVE - a positioning advantage the fighter tries to obtain just before the merge.*
5. AGGRESSIVENESS - how aggressively the fighter employs his aircraft weapon systems.
6. OFFENSIVE MANEUVERING - fighter's ability to optimize offensive position and achieve missile shots.
7. DEFENSIVE MANEUVERING - fighter's ability to maneuver while defensive and avoid being shot.
8. KEEPING SIGHT - pilot's awareness of position of wingmen and adversary aircraft.
9. ENERGY MANAGEMENT - optimizing airspeed while maneuvering.
10. MENTAL PLOT - fighter's mental picture of aircraft positioning while engaged.
11. SITUATIONAL AWARENESS - the total of ACM performance.
12. BUGOUT - technique used to disengage from ACM and arrive at a safe area.
13. WEAPON EMPLOYMENT - radar use in intercept and use of weapons while engaged in ACM.
14. VID TECHNIQUE - appropriate use of radar in the intercept.
15. VID COMMUNICATION - fighter-to-fighter and fighter-to-ground control, radar intercept communications.
16. UHF COMMUNICATION - fighter-to-fighter communication while engaged.
17. GAME PLAN - execution of tactical engagement plan.
18. MUTUAL SUPPORT - fighter's ability to protect and support wingmen.
19. DEBRIEF - participation in the fighter/adversary debriefing.
20. RECONSTRUCTION - ability to remember and reconstruct the ACM fight.
21. NUMBER OF KILLS - combination of measures 23 and 24.
22. NUMBER OF MISSILES LAUNCHED - self explanatory.
23. NUMBER OF VID KILLS - pre-merge* kills. (These are made prior to actual ACM, usually with forward-aspect missiles.)
24. NUMBER OF ENGAGED KILLS - post-merge* kills. (Those made during actual ACM, usually with heat seeking missiles.)
25. MEAN TIME TO FIRST KILL - calculated from 10-mile separation point of fighter and adversary aircraft.
26. NUMBER OF RADAR LOCKS - self explanatory.
27. RADAR LOCK MEAN RANGE - mean range at which radar lock obtained.
28. VISUAL TALLY-HO MEAN RANGE - mean range of initial sighting of adversary aircraft during intercepts.
29. VID MEAN RANGE - mean range of adversary aircraft identifications, i.e., "A-4."
30. NUMBER OF TIMES KILLED - self explanatory.
31. KILL-DIFFERENCE SCORE - measure 21 minus 30 - a composite score.
32. ENGAGED KILL DIFFERENCE SCORE - measure 24 minus 30 - a composite score.

* Merge point: The point at which the fighter and adversary aircraft first pass during the intercept.

## APPENDIX C

TABLE C-1. Subjective Measure Forward Selection, Analysis of Variance, Coefficient, *F* Values, and Model Summary Statistics.

Stepwise regression procedure for dependent variable (overall ACM grade)
Step 6, Variable V8 entered — Multiple *R* = .91; $R^2$ = 0.83; Adjusted $R^2$ (shrinkage) = .81

| | *df* | *SS* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Regression | 6 | 0.2658 | 0.0443 | 98.57 | 0.0001 |
| Error | 118 | 0.0530 | 0.0004 | | |
| Total | 124 | 0.3189 | | | |

| | B Value | *SE* | Type 11 *SS* | *F* | *p* |
|---|---|---|---|---|---|
| Intercept | 1.1111 | | | | |
| V3 | 0.0760 | 0.0124 | 0.0169 | 37.62 | 0.0001 |
| V6 | 0.0732 | 0.0111 | 0.0195 | 43.49 | 0.0001 |
| V8 | 0.0747 | 0.0143 | 0.0123 | 27.37 | 0.0001 |
| V9 | 0.0731 | 0.0136 | 0.0129 | 28.70 | 0.0001 |
| V11 | 0.0869 | 0.0087 | 0.0445 | 99.08 | 0.0001 |
| V18 | 0.0669 | 0.0077 | 0.0336 | 74.66 | 0.0001 |

Summary of stepwise regression procedure for dependent variable V1 - OAG

| Step | Variable Entered | Number in | Partial $r^2$ | Model $r^2$ |
|---|---|---|---|---|
| 1 | V11 Situational awareness | 1 | 0.4862 | 0.4862 |
| 2 | V6 Offensive maneuvering | 2 | 0.1132 | 0.5995 |
| 3 | V18 Mutual support | 3 | 0.0747 | 0.6742 |
| 4 | V3 Start/VID start | 4 | 0.0777 | 0.7519 |
| 5 | V9 Energy management | 5 | 0.0432 | 0.7951 |
| 6 | V8 Keeping sight/lookout | 6 | 0.0386 | 0.8337 |

TABLE C-2. Objective Measure Forward Selection, Analysis of Variance, Coefficients, *F* Values, and Model Summary Statistics.

Stepwise regression procedures for dependent variable (overall ACM grade)
Step 4, Variable V25 entered — Multiple *R* = .79; $R^2$ = 0.63; Adjusted $R^2$ (shrinkage) = .61

| | *df* | *SS* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Regression | 4 | 0.2008 | 0.0502 | 51.02 | 0.0001 |
| Error | 120 | 0.1181 | 0.0010 | | |
| Total | 124 | 0.3189 | | | |

| | B Value | *SE* | Type 11 *SS* | *F* | *p* |
|---|---|---|---|---|---|
| Intercept | 1.9973 | | | | |
| V23 | 0.0075 | 0.0020 | 0.0134 | 13.62 | 0.0003 |
| V24 | 0.0131 | 0.0016 | 0.0671 | 68.17 | 0.0001 |
| V25 | -0.0003 | 0.0001 | 0.0044 | 4.44 | 0.0371 |
| V30 | -0.0205 | 0.0026 | 0.0624 | 63.40 | 0.0001 |

Bounds on condition number: 1.5855, 42.7055

Summary of stepwise regression procedure for dependent variable V1 - OAG

| Step | Variable Entered | Number in | Partial $r^2$ | Model $r^2$ |
|---|---|---|---|---|
| 1 | V24 Engaged kills | 1 | 0.3273 | 0.3273 |
| 2 | V30 Number of times killed | 2 | 0.1919 | 0.5192 |
| 3 | V23 VID kills | 3 | 0.0968 | 0.6160 |
| 4 | V25 Mean time-to-first kill | 4 | 0.0137 | 0.6297 |

TABLE C-3. Subjective and Objective Measure Forward Selection, Analysis of Variance, coefficients, F Values, and Model Summary Statistics.

Stepwise regression procedure for dependent variable (overall ACM grade)
Step 4, Variable V18 entered — Multiple R = .89; $R^2$ = 0.78; Adjusted $R^2$ (shrinkage) = .78

| | df | SS | MS | F | p |
|---|---|---|---|---|---|
| Regression | 4 | 0.2503 | 0.0626 | 109.39 | 0.0001 |
| Error | 120 | 0.0686 | 0.0006 | | |
| Total | 124 | 0.3189 | | | |

| | B Value | SE | Type II SS | F | p |
|---|---|---|---|---|---|
| Intercept | 1.5826 | | | | |
| V11 | 0.0880 | 0.0101 | 0.0432 | 75.61 | 0.0001 |
| V9 | 0.0757 | 0.0154 | 0.0138 | 24.12 | 0.0001 |
| V18 | 0.0400 | 0.0088 | 0.0119 | 20.78 | 0.0001 |
| V31 | 0.0067 | 0.0008 | 0.0358 | 62.59 | 0.0001 |

Bounds on condition number: 1.5474, 41.0336

Summary of stepwise regression procedure for dependent variable V1 - OAG

| Step | Variable Entered | Number in | Partial $r^2$ | Model $r^2$ |
|---|---|---|---|---|
| 1 | V31 Kill difference score | 1 | 0.5724 | 0.5724 |
| 2 | V11 Situational awareness | 2 | 0.1356 | 0.7080 |
| 3 | V9 Energy management | 3 | 0.0395 | 0.7475 |
| 4 | V18 Mutual support | 4 | 0.0373 | 0.7848 |

# STANDARDISED TESTS FOR RESEARCH WITH ENVIRONMENTAL STRESSORS: THE AGARD STRES BATTERY

L.C. Boer
TNO Institute for Perception
3769 ZG Soesterberg
The Netherlands

E.W. Farmer
RAF Institute of Aviation Medicine
Farnborough, Hants GU14 6SZ
UK

and

G.F. Wilson
AAMRL/HEG, Wright-Patterson AFB
Ohio 45433
USA

## Summary

Assessment of human cognitive performance under stress is highly desirable, but is hampered by lack of standardization. Most tasks used in stress research are based on the paradigms of Human Performance Theory, that are frameworks for the manipulation of variables, not yardsticks for assessing stressor effects. In consequence, the results of two different studies cannot be compared directly. Even if the studies used the same paradigm, the particular task may have differed on such variables as condition of testing, number of trials, amount of training, or type of stimuli. Thus, for applied work there is a need for standardization.

The present paper describes the AGARD STRES battery (Standardized Tests for Research with Environmental Stressors) as proposed by AGARD AMP Working Group 12 "On Human Performance Assessment Methods". The battery consists of seven tasks based on widely-used paradigms. The seven tasks and a data-exchange format are described.

## Introduction

In wartime, military personnel will be exposed to many sources of stress. Sustained operations are very likely, and people will be required to work at night-time hours when natural alertness is at its lowest level. Moreover, sleep loss and fatigue will accumulate. Working for long periods in complete isolation may occur. Protective drugs causing blurred vision and reduced alertness may be prescribed, as well as protective garments and face masks causing such discomforts as skin irritations, thermal stress, reduced dexterity, social isolation, and the inability to eat food. Threat to life and limb will be present, and anxiety and worry may be manifest in others as well.

There is an obvious interest in finding out what effects these environmental stressors have on human capability, and what levels of performance can be maintained under stress. The flying task has received particular attention, because of the potentially disastrous consequences of stress-induced performance failures.

Assessment of stressor effects is not an easy task. The main problems are: (a) It may be impossible to determine "the" stress vulnerability of a particular task, because a task may be sensitive to one stressor but unaffected by another. Moreover, stressors may interact (their combined effect being larger than the sum of the effects in isolation) or counteract (their combined effect being smaller than the sum of their isolated effects). A lot of research using multiple stressors may thus be necessary. (b) It may be impossible to make predictions on the level of the individual subject, because subjects may vary considerably in their sensitivity to a stressor. Results will thus hold for the average person, that is, the mean of a group. (c) Choice of task may be difficult because many different performance tasks are useful to assess stressor effects, and there is no standardization.

Most tasks used in applied research are based on the paradigms of Human Performance Theory. These paradigms are developed as frameworks to investigate how manipulation of variables affects information processing -not as yardsticks for assessing stressor effects. In consequence, the results of two different studies cannot be compared directly. Even if the studies used the same paradigm, the particular task may have differed on such variables as condition of testing, number of trials, amount of training, or type of stimuli. Thus, for applied work there is a need for standardization.

Confronted with this need for standardization, a number of researchers in the NATO member countries met with the intention of introducing a standardized approach to performance testing. The "Aachen Academic Group" held a series of meetings sponsored by the USAF and the European Community. This endeavour was continued under the auspices of AGARD's Aerospace Medical Panel as Working Group 12 "On Human Performance Assessment Methods". The mission of the Working Group was to construct a standardized test battery. Steps to be taken were (a) selection of tests, (b) review of previous literature on each test, (c) standardization of parameters, and (d) specifications for a data-base.

The present paper is the first report on the AGARD STRES battery. STRES is an acronym for Standardized Tasks for Research with Environmental Stressors. The full report will appear as AGARDograph 308, containing descriptions sufficiently detailed to enable a programmer to implement the battery on his own system. Moreover, a general software package for running the STRES battery may become available soon. As a further introduction of the battery, other activities are planned, such as an AGARD Lecture Series (#163) to be held in Canada, the Netherlands, and Italy in June 1989.

Before describing the tasks of the STRES battery, the standardization problem is discussed, and the application areas of Human Performance tasks and their theoretical background are reviewed. Readers who are interested in the tasks of the battery could skip the next two, three pages, and resume reading at the section called Test Specifications.

**Performance Testing and the Need for Standardization**

Psychological tests are traditionally presented in a completely standardized form. Publication is usually preceded by a lengthy development phase. The performance tests used in stress research, however, have a quite different history. Often, they are borrowed from techniques reported in the theoretical literature on human cognition. The strength of these techniques is the possibility of manipulation of all kinds of task parameters. They are research tools. Consequently, no standard protocols are available, and it is not surprising that, when applying these "paradigms" as tools for the assessment of stressor effects, researchers construct their own versions of the test that, although conforming to the general paradigm, differ considerably in detail.

Sanders, Haygood, Schroiff and Wauschkuhn's (1) survey of performance test batteries, and the discussions of performance researchers comprising the "Aachen Academic Group", indicated that the bottleneck of a standardization programme is not the selection of tests. There appears to be a surprising consensus in the selection of tests for different batteries. Batteries usually have a Sternberg (2, 3) memory search task; they usually have a tracking task, etc. The problem is rather the unlimited variation in all sorts of parameters within one and the same paradigm. The Aachen Group concluded that a standardized battery based on a core of commonly used performance tests could easily be compiled, and that the main task would be the standardization of parameters. The establishment of a normative data base, comparable to that available for intelligence and personality tests, could only then be established.

Working Group 12 of the AGARD Aerospace Medical Panel was formed to achieve the objective of a standardized battery. The seven most common paradigms were selected as the basis of the AGARD STRES Battery, each with some evidence of psychometric soundness. The battery is intended to become a yardstick against which the effects of stressors can be assessed. Moreover, it provides the applied researcher with a solid core of well-accepted and well-defined performance tests. Because the accumulation of a database is essential for the current project, specifications for a data storage format and a preliminary database were established.

The benefits of this standardization programme include the opportunity to apply both "narrow-band" and "broad-band" strategies (4) to stress research. The narrow-band approach involves examination of the effects of a variety of stressors on performance of a single task, and permits generalizations concerning the effects of stressors; the broad-band approach, in which the effects of a single stressor on various tasks is investigated, helps to reveal subtle but important differences between stressors (see 5 for a similar distinction). The data base will also permit examination of the role of subject variables such as age, sex, and occupation.

**Applications of Human Performance Testing**

There are two broad classes of purpose for a battery of performance tests. Firstly, the battery can be used to evaluate the effects of environmental stressors; or, secondly, to assess the information-processing abilities of individuals. When the purpose is to evaluate stressor effects, emphasis is placed upon comparison of the performance of groups of subjects tested with and without unfavourable conditions such as sleep loss and fatigue; monotony and boredom; illnesses; toxic fumes; hypoxia; alcohol and other drugs; and the wearing of cumbersome garments. The ultimate goal is to assess the extent to which a particular stressor influences performance in real-life situations. When, by contrast, the purpose is the assessment of abilities, interest lies in differences between individuals. This application is comparable to classical test psychology. The individual's score is used as a measure of information-processing ability relative to that of other individuals.

The AGARD STRES Battery is concerned primarily with stressor assessment, that is, the accent is on performance under unfavourable environmental conditions (fatigue, boredom, illness, toxic agents, cumbersome equipment, etc.). The requirements of stressor assessment differ in some respects from those of ability assessment. To assess individuals, test measures should ideally be relatively insensitive to variations in environmental conditions but sensitive to individual differences.

To assess stressor effects, the opposite is true: Performance should fluctuate markedly when environmental conditions change, but the variance due to individual differences should be as small as possible (see Figure 1).

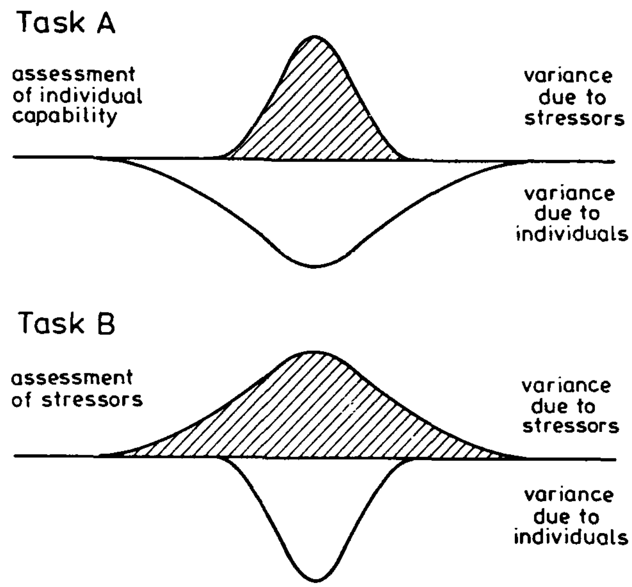



Figure 1. Differential sensitivity to stressors and individual differences. (Task A is more sensitive to individual differences; Task B is more sensitive to stressors.)

In practice, however, a test may be found to be sensitive both to stressor effects and to individual differences, and for this reason the potential application of the STRES battery to personnel selection will not be ignored.

**Human Performance Theory: Scope and Limitations**

The STRES Battery is not dependent upon a specific theoretical standpoint. Nevertheless, it is instructive to consider the general nature of models of human performance, the mental processes that commonly used performance tests purport to measure, and the ways in which these tests differ from many real-life activities and tasks.

The aim of Human Performance Theory (HPT) is to search for lawful relations between task variables and performance. This has led to the development of a large number of information-processing models. The central assumption underlying most models is that human cognition functions with the characteristics of a single information-processing system equipped with memory stores, or with the characteristics of an ensemble of such systems each with its own functional significance. This so-called computer analogy incorporates the notion of limited capacity, which implies both that mental processes are time-consuming and that the time required increases with complexity. Thus "mental chronometry", in which mental processes are investigated by dissection of reaction time (RT), is one of the most important tools of HPT.

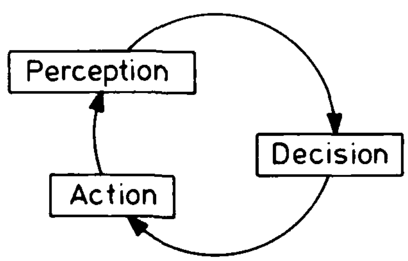



Figure 2. The Perception-Decision-Action model.

A very general information-processing model is that of the Perception-Decision-Action cycle shown in Figure 2. Perception and action are the input and output functions, respectively, with decision as the intervening process. Figure 3, which shows the various stages of the reaction-time process in addition to some of the task variables affecting these stages, can be considered a more specific elaboration of the model.

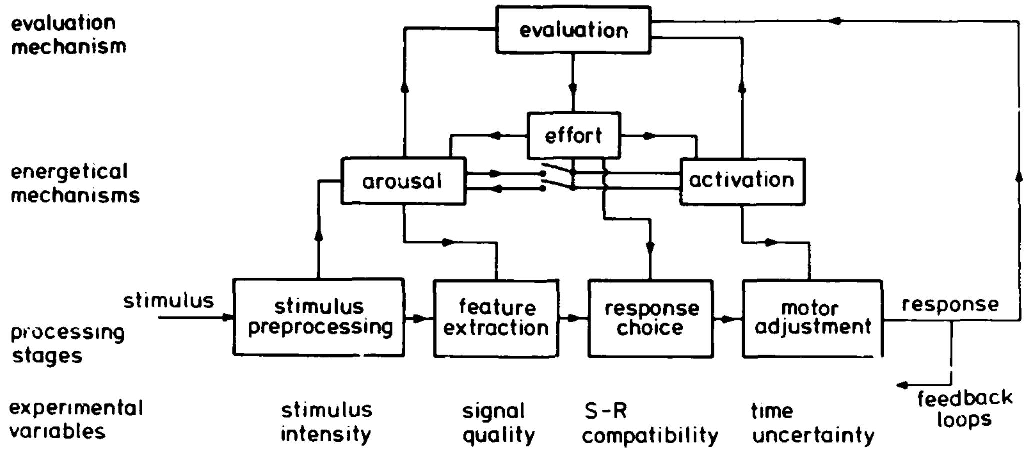



Figure 3. Energy and structure. The model shows the structure of the reaction process (bottom line), and energetical supply to the structural elements. (Adapted from Sanders, 6.)

The boxes in the bottom line of Figure 3 represent the structural properties of information processing. The model also takes into account the dimension of energetical supply. The supply to perceptual processes is called arousal, and that to motor-related structures is called activation. The concept of energetical supply, or amount of mental resources available to the information processing structures, is very important in the present context of stress research, because stressors affect the supply of energetical resources.

Normally, resources for adequate task performance are allocated to processing structures with little conscious effort. Stressors, however, may hinder the supply of resources, either by reducing the total amount of energy available, or by directing the flow of energy to activities unrelated to, or even detrimental to, adequate task performance. Energy reduction has been postulated to occur under conditions of fatigue, boredom, and sleepiness ("worn out"); energy diversion and task interference under conditions related to anxiety and worry ("worried out"). As a consequence, tasks are not always provided with the necessary resources, and information-processing performance will suffer. The extent of performance deterioration indicates the severity of the stressor effects.

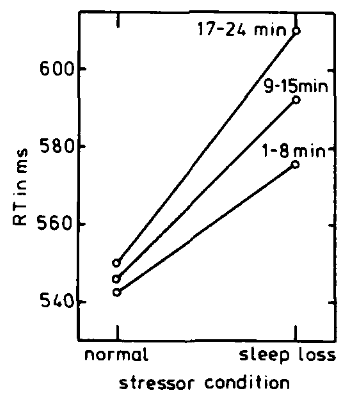



Figure 4. Performance as a function of stressor condition, with time on task as the parameter. (Sleep loss and fatigue have the effect of slowing down RT. Source: 43.)

It is important to recognize that all such inferences about stressors have an indirect quality. For example, fatigue results in a deterioration of human cognitive performance (see Figure 4 for an illustration). It is therefore quite legitimate to suggest that performance tasks can measure fatigue. However, one should bear in mind that mental performance may also be affected by factors such as other stressors, differences in individual capability, and amount of practice. A "blind" interpretation of a deterioration in performance can be very misleading, and a full knowledge of the situation is required. For this reason, investigators try to manipulate only the stressor of interest while eliminating any confounding due to the presence of other stressors, due to individual differences, and due to practice. Interpretation of performance is possible only in such controlled environments.

Clearly, the tightly-constrained paradigms of HPT sample only a subset of human behaviour. This is apparent even if only the dimension of complexity of processing is considered. Complexity of information processing is determined by the nature of the stimuli, the rule by which stimuli are mapped to responses, and the type of response required. HPT paradigms use only highly-structured information-processing tasks. Stimuli are well-defined units such as letters, words, or tones; responses are key-press reactions or simple vocal utterances; and stimulus-to-response (S-R) mappings are unambiguously specified. Moreover, the tasks have well-defined starting and end points, and exclude the ambiguity encountered in real-life activities, in which the individual must sometimes determine the exact nature of the situation before deciding what, if any, action is required. Unusual and unexpected events are unlikely in HPT tasks.

Real-life stimuli may be extremely complex. They may comprise many different elements, perhaps requiring temporal integration over prolonged periods of time; they may be hidden or masked by other meaningful stimulus patterns; and they may occur unexpectedly. At the highest levels of complexity, the classification of stimuli may represent a source of contention even among experts, as for example in a difficult medical diagnosis, or an early assessment of a political or economical emergency.

Responses and S-R mapping rules may also be more complex in real life than in HPT tasks. Furthermore, these tasks typically present a repetitive succession of very similar but discrete S-R cycles. A real-life task, on the other hand, may comprise a single S-R cycle. Moreover, real-life tasks may lack well-defined starting or end points, and may have cumulative aspects in which task difficulty depends on past performance. In most performance tests, cumulative effects are limited to fatigue and practice effects.

In summary, it is apparent that the focus of HPT is on the basic mechanisms of cognitive information processing. The conclusions of stressor studies should be viewed in this light. The HPT test depends upon tightly constrained domains of stimuli and responses, and samples relatively low-level behavioural tasks. It is therefore obviously most relevant to similar well-defined real-life tasks. Such real-life tasks can be identified, for example, within the activities of the aircraft pilot. When controlling the attitude of an aircraft, the pilot must extract signals concerning the position of the horizon, and make relatively simple manual corrections. Other real-life tasks, however, bear little or no obvious relationship to the mental processes measured by performance batteries based on HPT. For example, the complex decision processes required of the military commander are a different level of processing, and are not well represented by performance tests requiring specific responses to well-defined stimuli.

**Psychometrics**

Any psychological test must measure what it purports to measure, do so consistently, and be capable of detecting the effects of the environment or of individual differences in ability. In other words, it must exhibit the psychometric properties of validity, reliability, and sensitivity.

Construct validity indicates the extent to which performance is consistent with the theoretical notions concerning the nature of the mental process that the test is designed to measure. Concurrent validity, which refers to the relationship between the attributes measured by the tasks and other measures of the same attribute, is also relevant to the STRES Battery. Finally, external validity, which refers to the correlation between the individual's test score and his/her performance in real-life tasks, is important too.

Performance tends to become more stable after practice. The reliability of a performance task is, therefore, increased as more practice is provided. The final specification of each STRES task includes a standard training schedule to ensure that most of the effects of practice are eliminated prior to the experimental phase. It is strongly recommended that this training schedule be adopted. The available evidence suggests that, after training, STRES task scores will have an acceptable level of reliability.

The existing evidence of reliability, validity and sensitivity for the STRES battery tasks is reviewed in AGARDograph 308. The following approaches to formal validation studies were identified by the Working Group and seem to be desirable:

(a) Use of factor analysis to relate the measurements of the battery to a well-established ability factor space, such as that formed by Cattell's Comprehensive Ability Factors.

(b) Assessment of construct validity by administering the tests to various occupational groups. It can be predicted, for example, that a group of successful pilots will score more highly than a group of radio operators on the Spatial Processing task.

(c) Assessment of external validity, or the degree to which decrements in test performance reflect decrements in real-life performances. For example, the user must be able to infer the operational consequences of a particular pattern of decrement in test scores under sleep loss.

(d) Assessment of cross-cultural validity. It must be ensured that performance on the tasks is not affected by cultural differences. For example, the Grammatical Reasoning Test, as described by Baddeley (32) would be unsuitable for use in German, because of the avoidance of the passive voice in that language.

## Test Specifications

The survey conducted by Sanders et al. (1) was used initially to identify tasks based on the paradigms of Human Performance Theory that were in common use and that together would provide measures of a wide range of mental processes. Individual tests were then selected on the basis of the following criteria:

(a) Preliminary evidence of reliability, validity, and sensitivity.

(b) Documented history of application to assessment of a range of stressors.

(c) Short duration (maximum of three minutes per trial block).

(d) Language-independence.

(e) Sound basis in Human Performance Theory.

(f) Ability to be implemented on simple and easily-available computer systems.

The following seven tests, comprising a total test time of 50 min, were selected on the basis of these criteria.

**Reaction time.** Several reaction time tasks satisfy the criteria listed above. Since the duration of a single S-R event is typically very short, a large number of trials can be accommodated within a three-minute period. The task selected was based on that appearing in the TNO Taskomat Battery (7), since it provides separate measures of the stages comprising the reaction process.

**Mathematical processing.** Numerical ability has repeatedly been identified as a factor in factor-analytic studies of skilled performance. Several mathematical processing tasks exist, but most require a numerical response. The Mathematical Processing task from the USAF Criterion Task Set (CTS, 8) and the Unified Tri-Services Cognitive Performance Assessment Battery (UTC-PAB, 9) was chosen since its two-choice response is more suitable for computerized presentation. This task measures the ability to manipulate arithmetical information, and so places demands upon working memory.

**Memory search.** The Sternberg memory search paradigm was selected because of its popularity in applied performance studies, and its possibility to separate processing times associated with memory search and comparison from other processing times.

**Spatial processing.** Spatial processing tests exist in a variety of forms, some requiring complex hardware. The CTS/UTC-PAB version, which taps visuo-spatial short term memory by requiring the subject to imagine rotations, was selected because of the well-documented history of application of this general technique, and its ability to be administered using relatively simple hardware.

**Unstable tracking.** Tracking places demands primarily upon motor-related resources. Of the many tracking tests available, the CTS/UTC-PAB version was selected because of its previous application to stress research, and its sound theoretical basis.

**Grammatical reasoning.** Though some researchers have argued that mathematical and verbal reasoning tasks sample the same resource, it has been reported that performance on these two types of test can be differentially affected by some stressors. Both types of test were therefore included in the present battery. The STRES grammatical reasoning task requires the manipulation and comparison of grammatical information. It was adapted from a task described by Baddeley (33), which has been used extensively to measure stressor effects.

**Dual-task performance.** Division of attention between task components is an important element of many practical tasks such as flying, and there is evidence that the allocation of mental resources is affected by stress. It was therefore considered essential to include in the battery a measure of dual-task performance. Since dual-task performance can be interpreted only in the light of performance on each task in isolation, the total administration time of the battery was reduced by combining two of the tasks already included in the battery. Tracking and memory search were selected because of their relevance to continuous control tasks, such as flying, in which there are periodic demands upon working memory.

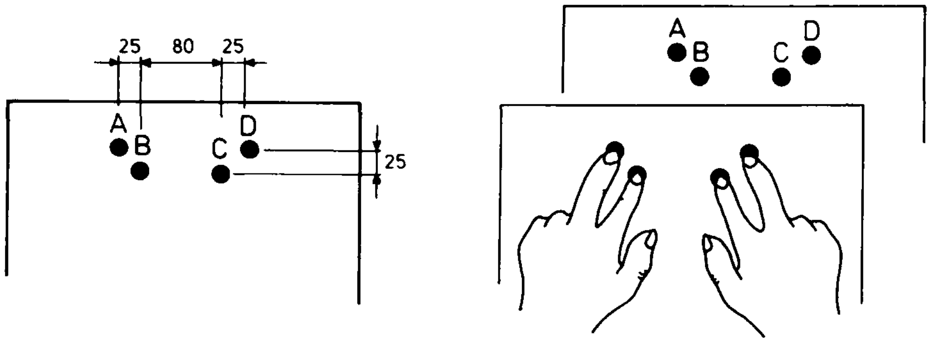



Figure 5. Response key configuration (distances in mm).

The seven tests have several features in common. For example, they all use a computer monitor to display the stimuli; and, with the exception of the Tracking task, they all use the four-key response panel shown in Figure 5. Further information on these common features and the duration of each task is provided after the description of the individual tests.

**Reaction Time Task**

The Reaction Time (RT) task can be used to test the effects of stressors on the separate stages comprising the reaction process. Basic RT is measured in a first and a last (6th) block; the intervening blocks (2-5) present more complex trials, each block loading a specific stage of the reaction process, as indicated in Figure 6. The RT differences between complicated and basic blocks give specific information about the effect of loading four specific stages.

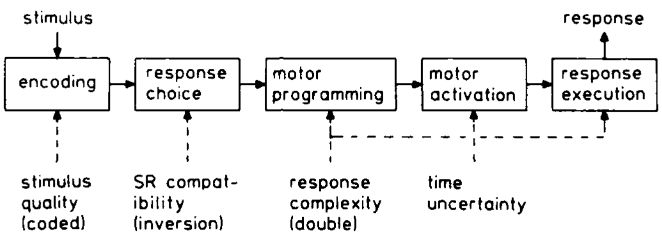



Figure 6. Stages of the reaction process, and the effects of task variables.

Digits are presented on a computer monitor, one every 2 s. The subject reacts to each digit by pressing the appropriate key on the response panel. S-R mapping is based on (a) position of the digit, either left or right, and (b) identity of the digit. The left hand (resting on keys A and B) is used in response to digits presented on the left side of the screen, and the right hand (keys C and D) is used in response to digits presented on the right side of the screen. Within the hand indicated by the digit's position, a further specification of the finger to be used is given by the digit's identity. Digits 2 and 3 specify a left finger (middle finger, or key A, for the left hand; and index finger, or key C for the right hand); digits 4 and 5 specify a right finger (index, or key B, for the left hand; and middle finger, or key D for the right hand).

Manipulated across trial blocks are the following task variables in the following, fixed, order: stimulus quality (degraded as shown in Figure 7), time uncertainty (stimulus-to-stimulus interval not 2 s, but a random choice between 2000 and 10,000 ms), response complexity (a triad of key presses, such as ABA or BAB, is required instead of a single keypress response), and compatibility of S-R mapping (the right hand is used in response to digits presented on the left, and vice versa).

<u>Background</u>

The idea that the process between stimulus presentation and overt reaction contains a number of discrete steps or stages dates back at least to Donders (10), who tried to estimate the duration of decision processes by subtracting simple (non-choice) reaction times from choice reaction times. His basic ideas were revived in the 1960s. For example, the second Attention and Performance symposium was called the Donders Centenary Symposium on Reaction Time. One of the significant events was Sternberg's presentation of "Extensions of Donders' Method" (11), which helped to introduce the Additive Factor Method. This method was based on the premise that processing stages can be identified by investigating the relation between different task variables rather than between different tasks as proposed by Donders.

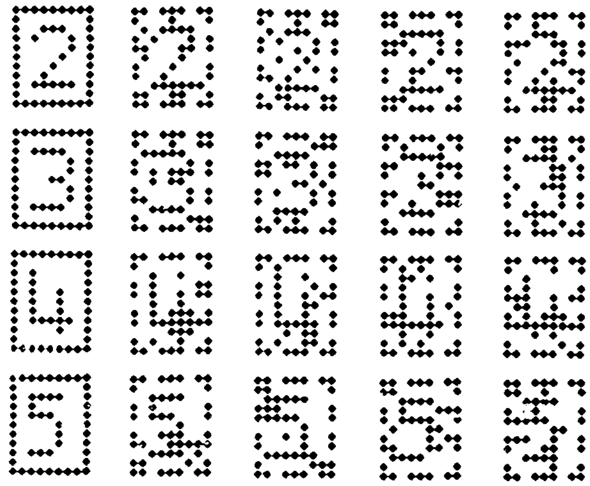


Figure 7. Normal and degraded stimuli used in the Reaction Time task. Stimuli are surrounded by a rectangular frame of 57 x 46 mm. Degraded versions of each digit are created by moving 10 elements from the frame towards the figure. Each element comprises two triangles, situated side by side with one pointing to the left and the other pointing the right to form a diamond shape. The grid on which the triangles are placed is the same as that used for normal presentation of text.

The Additive Factor Method became an influential research method, and many subsequent studies on the effects of task variables were conducted. The model of Figure 6 (cf. also Figure 3) is a summary of this research.

**Mathematical Processing Task**

The purpose of this mental arithmetic task is to place demands upon the processing resources associated with working memory. The test requires subjects to perform two arithmetical operations, addition and/or subtraction, on a set of three single-digit numbers, and to determine whether the answer is greater than or less than five. Problems are presented in the centre of the monitor screen in a horizontal format (e.g. 5 + 3 - 4 =); the subject is instructed to solve the problem working from left to right, and to press the key marked ">" or "<". The problem disappears when a key is pressed, and after a random interval of between 3000 and 5000 ms a new problem appears on the screen.

Background

The present test, developed by Shingledecker (8), requires the execution of two mathematical operations (addition and/or subtraction) within a given problem. Chiles, Alluisi, and Adams (27) developed an initial version of the mathematical processing task, requiring both addition and subtraction, for use in the assessment of mental workload. This task was included in the Multiple Task Performance Battery (MTPB) with other cognitive tasks such as auditory vigilance, warning lights, meter monitoring, problem solving, choice reaction time, tracking, and pattern discrimination; it was used in multi-task studies to examine subjects' time-sharing ability (e.g., 12, 13, 14, 15).

Chiles, Jennings, and Alluisi (16), using multi-operation problems, reported a pattern of dual-task interference consistent with the notion that mathematical processing taps working memory resources. Performance on an arithmetic task was poorer with a concurrent code lock solving task than with a concurrent manual tracking task that placed demands primarily upon response-based resources. Research with single-digit addition problems (e.g., 17) has supported the hypothesis that adults solve simple addition problems by recourse to information stored in long term memory. Moreover, research with multi-digit addition problems (e.g., 18) has shown that complex mathematical problems are solved in a series of elementary steps requiring storage of intermediate results in working memory.

**Memory Search Task**

This task examines the ability to search items held in memory for the presence of a "probe" item. A set of letters (the "memory set") is presented on the display. When the subject has memorized the memory set, he/she presses a key, and a series of single letters (the "probe letters") the correct response would be "no". The probe is erased as soon as a response key is pressed, and a new probe appears one at the time. The subject has to indicate, by pressing an appropriate key, whether the probe letter is a member of the memory set. For example, if the memory set were G, X, T, L and the probe letter were T, then the correct response would be "yes"; if the probe letter were D, then the correct response would be "no". The probe is erased as soon as a response key is pressed, and a new probe appears one second later. Each trial block has a duration of three min. The test consists of two such 3-min blocks, Block 1 using a memory set size of two letters, and Block 2 a memory set size of four letters. The difference in average performance between the two blocks is a measure of the speed of memory search and comparison.

Background

RT increases linearly as a function of memory set size. The slope of the RT/memory set function is assumed to reflect the speed of the processes involved with memory search and comparison, while the intercept with the RT axis is assumed to reflect the speed of all other reaction processes (2). Sternberg (3) further demonstrated that degradation of the probe affects only the intercept of the RT function, indicating that this manipulation affected some stage other than that of memory search-and-comparison. This stage was called "recognition" (or "encoding" - see Figure 6). It was concluded that the probe was "cleaned up" prior to memory search, increasing RT by a constant amount regardless of memory set size.

The effects of practice are noteworthy. If the same fixed memory set is used over many days, then the RT function becomes flatter and negatively accelerated (19, 20). There is evidence that subjects develop a content-addressable search strategy (21), and that processing becomes automatic rather than controlled (22, 23). If, as is recommended for the present STRES version of the task, the memory sets are changed after each block, so that stimuli are not consistently associated with particular responses, then extended practice affects the intercept but not the slope (24).

**Spatial Processing Task**

This task is designed to examine the subject's ability to make spatial transformations (see Lohman's, 5, survey of the correlation literature on spatial ability). On each trial, a standard four-bar histogram is presented for three seconds, followed after a one-second interval by a similar but rotated test histogram (see Figure 8). The subject must determine whether the test histogram is identical to the standard histogram, regardless of the orientational difference of 90 degrees or 270 degrees, and respond "same" or "different" by pressing the appropriate response key. Responding erases the stimulus and after one second a new pair of histograms is presented. This continues for three minutes.

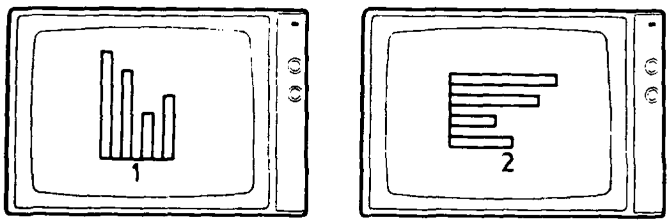



Figure 8. Sample stimulus of the Spatial Processing task. (The unit of the bars is a 8.5x5.0 mm--height x width--rectangle. Bars consist of 1-6 of these units.)

Background

This task is adapted from the spatial processing task used in the CTS (8) that is, in turn, derived from an earlier task devised by Fitts et al. (26) and later used by Chiles, Alluisi, and Adams (27). Fitts and his colleagues presented a single histogram as a standard, followed by six rows of eight simultaneously presented test stimuli. The subject's task was to select from each row the test stimulus that was identical to the standard.

**Unstable Tracking Task**

This task tests motor-related processing resources used in the execution of continuous manual control responses. The subjects see a fixed target in the centre of the monitor screen. They manipulate a joystick in an attempt to maintain the position of a horizontally-moving cursor on the target. The system is inherently unstable: Operator input introduces error that is magnified such that it becomes increasingly necessary to respond to the velocity as well as the position of the cursor. The dynamics of the task are analogous to those of balancing a stick on one's fingertip (28): As soon as an error from the vertical is introduced, the stick will begin to fall, its rate of fall increasing as it falls.

The cursor, a rectangle 5 mm wide and 10 mm high, moves horizontally within a display area 95 mm in width. Subjects are required to maintain the position of the cursor on the central target, avoiding control losses in which the cursor reaches the edge of the screen. Subjects are given 10 s to gain control of the cursor before the three-min data collection period begins. If the subject loses control and the cursor reaches the boundary line, a control loss is recorded, and the cursor is automatically re-positioned on the target, and the subject continues tracking. This continues for three minutes. Average tracking error (rad/s) and number of control failures are recorded.

The unstable plant dynamics of the STRES Unstable Tracking task are a first-order divergent element of the form:

$$P(s) = \frac{\lambda \exp^{-ts}}{s - \lambda}$$

where P = ratio of system output to input
s = Laplace operator (system response is a function of frequency)
λ = instability of the cursor = 1/T, where T (in seconds) is divergent time constant
$\exp^{-ts}$ = additional phase lag produced by time delay t.

For a system with a screen-refresh rate of 50 Hz, the position of the cursor is determined by the following relationship:

$$\text{New\_pos} = \frac{(2R + \lambda)}{(2R - \lambda)} \text{old\_pos} + \frac{\lambda * \text{Gain}}{(2R - \lambda)} (\text{stick\_input} + \text{last\_stick\_input}),$$

which, with

R = Refresh Rate = 50 Hz,
λ = 2 (because λ = 1/T, T = 0.5 seconds) and
Gain = 4,

reduces to

$$\text{New\_pos} = 1.0408 * \text{old\_pos} + 0.0816 * (\text{stick\_input} + \text{last\_stick\_input}).$$

Background

This task was developed by Jex, McDonnell, and Phatak (29). It was inspired by analytical treatment of aircraft handling qualities, such as Ashkenas and McRuer's (30) work on just-controllable aircraft short-period static instability and its strong relationship with operator (pilot) effective time delay. Ashkenas and McRuer corresponding increases in the operator's internal delay in processing and responding to the disturbance. Subsequently, it was reported that control loss occurred at the same static instability level for three test pilots (31). These findings resulted in a more extensive investigation of the dynamics of manual control behavior, and provided the impetus for the development of a reliable, internally valid control task for applied research. Figure 9 shows an example of a describing function with effective time delay and gain as the parameters.

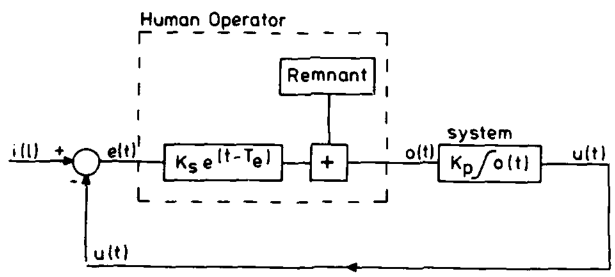



Figure 9. The Quasilinear Crossover Model. The subject's output at time t, o(t), is a function of the gain parameter, Ks, and time delay, $T_e$.

The precise parameters of the Unstable Tracking task were determined empirically during Shingledecker's (8) test development phase.

**Grammatical Reasoning Task**

This task, derived from that described by Baddeley (32), addresses the ability to manipulate grammatical information, placing demands primarily upon working memory. The task is basically a sentence verification task. On each trial, two sentences (actively and positively phrased) are presented, together with a row of three adjacent symbols. The sentences describe the order of the symbols, and the subject must compare the veracity of the description of the order of symbols contained in the sentences. A sample stimulus may consist of the sentences # AFTER * and # BEFORE & and the symbol row *#&. If both sentences are true descriptions of the actual order of symbols (as is the case in the example) or if both sentences are false descriptions of the actual order, a "same" response is required. If the sentences have opposite truth values (one "false", the other "true") a "different" response is required. Pressing a response key erases the stimulus and initiates, after a one-s interval, a new stimulus.

Background

Baddeley's (32) grammatical reasoning task was inspired by findings reported by Slobin (33) and Wason (34). In Baddeley's task, a statement describing the order of letters A and B was accompanied by the letter pair AB or BA (eg B is not followed by A - BA); subjects were required to indicate whether or not the statement correctly described the letter pair. Baddeley and Hitch (35) and Hitch and Baddeley (36) showed that a concurrent memory load of six letters slowed verbal reasoning performance but had no effect upon accuracy. Thus, it appeared that the short-term memory store and the system responsible for reasoning were at least partially overlapping. There is little doubt that verbal reasoning places demands upon central resources. Moreover, verbal reasoning places some additional demands on a subsystem called the articulatory loop. Farmer, Berman and Fletcher (37) found that articulatory suppression interferes with verbal reasoning.

The STRES version substitutes the symbols used by Clark and Chase (38) for the letters A and B, following Shingledecker's (8) modifications, and further departs from the original technique by abandoning the use of the passive voice, which is infrequent in some countries (e.g. Germany) and might therefore be responsible for cultural differences in test performance. In an attempt to redress the reduction in difficulty caused by elimination of passively phrased sentences, two instead of one sentences specifying the order of three symbols are presented on each trial.

**Dual Task: Tracking with Concurrent Memory Search**

This combination of the Unstable Tracking and Memory Search tasks measures the ability to divide attention between two activities. There are, just as in the Memory Search task described earlier, two blocks of three minutes. The first block is devoted to tracking while judging probe letters with a memory set of two, and the second block is devoted to tracking while judging probe letters with a memory set of four. Subjects are instructed to give the tracking and memory search tasks equal priority.

The task is administered only after the subjects have completed the component tasks in isolation. The dual task then begins with presentation of the memory set. The subject presses a response key to indicate successful memorization of the memory set. Then, a 10-s warm-up period of Unstable Tracking begins. The memory set remains on the screen for the first nine seconds of this period. After the tracking warm-up, the first probe letter is presented and the three-min dual task period begins. Probe letters are presented directly above the centre of the tracking target, with a vertical separation of seven mm.

Background

Shingledecker, Acton, and Crabtree (39) combined the tapping task of Michon (40) with tracking and memory search. The Michon tapping task interfered with tracking, but had no effect upon memory search performance. Since the Michon task is assumed primarily to tap resources associated with response timing, this pattern of dual-task interferences supports the hypothesis that the burden of tracking is on resources associated with response processing.

The STRES task combination employs the memory search configuration most likely to interfere with tracking because both tasks involve visual input and manual output, and hence there is competition for input and output resources (see 41). Moreover, task-hemispheric integrity is low in the present configuration: Subjects respond to the verbal memory task with the nonpreferred hand that is controlled by the cerebral hemisphere specialized for spatial processing, while they respond to the spatial tracking task with the preferred hand that is controlled by the hemisphere specialized for verbal processing (cf. 42).

## Conditions of Testing

It is important to standardize not only the tasks used to assess stressor effects, but the conditions of testing as well. The recommendations presented below should be followed as closely as possible. Deviations, where necessary, should be recorded carefully with the experimental data.

### Order of tasks and data

After subject information has been entered, the tasks are presented in the order indicated earlier (Reaction Time, Mathematical Processing, ..., Dual Task). At the completion of each individual task, condition information and performance data are stored on computer disk. Performance data consist of summary data over the relevant conditions of the test (eg, average reaction time, percentage of errors, number of response failures). These data can be entered into the database. Data for the individual trial are also stored (stimulus, response, reaction time) so as to give researchers the possibility to perform additional analyces on their data.

### Stimulus display

Display information is presented as bright elements on a dark background; the ratio of display element to background luminance should be between 7:1 and 12:1. Alphanumeric characters should subtend a vertical visual angle of 15-20 min of arc, which, at the recommended viewing distance of 0.6 m, corresponds to a character height of 2.6-3.5 mm. A "normal" computer display will usually be sufficient to meet these specifications. Because of the test battery's dependence upon presentation of visual material, it must be ensured that subjects have normal or corrected-to-normal vision.

### Response devices

To run the tests comprising the STRES battery, a four-response key panel and a joystick are required.

Depression of a response key should cause RT to be recorded to the nearest ms. Non-latching, push-to-make switches should be used, with a travel of three mm and an actuating force of 0.3-0.35 N, equivalent to application of a weight of 300-350g. The response key configuration and finger assignment were already shown in Figure 5. Most tasks need only two keys, C and D, but lefthanded subjects should use keys A and B instead. Only the Reaction Time task needs all four keys.

If no response key device is available, use of the computer keyboard would be acceptable as an emergency solution. For example, keyboard keys W,D,J,I could be used instead of the reponse panel keys A,B,C,D. If this solution has to be chosen it should be clearly recorded in the data file.

In the tracking task, the subject moves the joystick left or right to control the movement of a cursor on the screen of the computer monitor. The joystick lever and potentiometer should satisfy the following requirements:

(a) The range of movement of the lever should be 30 degrees left and right from the vertical position.

(b) The friction of the moving parts should not exceed 50 g, and should be constant over the range of travel.

(c) The relationship between angular rotation of the joystick and lateral movement of the cursor should be linear for the entire range of travel.

(d) Analogue-to-digital conversion of joystick potentiometer values should be conducted at eight-bit resolution or better. In other words, rotation of the joystick should produce at least 256 discrete values.

### Testing environment

External disturbances should be minimized during administration of the battery. If subjects are tested in groups, the test room should ideally be partitioned into separate workstations.

The position of the computer monitor relative to windows and sources of artificial light should be selected carefully. It is undesirable that the monitor appears as a silhouette against a bright window or, if the window is behind the subject, that the sun illuminates the screen directly thereby reducing the contrast ratio effectively to unity. It is also undesirable that the subject is distracted by reflections on the screen.

The surface of the screen should be perpendicular to the subject's line of sight, and should be located 0.6 m from the eye; smaller or greater distances are acceptable if the size of individual characters is adjusted to maintain the visual angle within the specified range. The seat height should be about 0.45 m, and the height of the upper surface of the response console about 0.75 m.

## Training

Performance improves significantly as a task is practised. To avoid confounding between the effects of stressors and of task learning, the latter must be minimized. Ideally, subjects should practise the task until their performance is stable. The tasks comprising the STRES battery differ in the amount of practice necessary to achieve reasonable stability. A standard training schedule is specified, usually consisting of ten three-min blocks per task. An abridged schedule of mostly two three-min blocks per task is also specified, and may be required if practical constraints limit the time available for testing.

Because some effect of training is likely to be observed during the experimental phase (especially if the abridged schedule is adopted) particular attention must be paid to balancing the order of test conditions. For example, in a normal versus stressor condition test design, training effects can be controlled by running half of the subject group through the normal condition first, while the other subjects complete the stressor condition first. This balances or controls the effects of practice.

## Task duration

The complete battery is administered within 50 minutes. Five of the tasks need only four min each: one min to initiate a task, and three min to run a trial block. The Memory Search task needs twice as long, because it consists of two blocks. The Reaction Time task needs 15 min: 12 min to run its six blocks of two min each, and three min to initialize.

The standard training schedule takes five hours. The Reaction Time task needs 90 min of training; the other tasks on the average 35 min.

When testing the effect of a stressor, it is desirable to adhere to the duration specified for each trial block. However, if the effects of a stressor are unlikely to become apparent within this limited time period, a multiple of the specified value may be used (see Figure 4 for a research example).

## Data Exchange

An essential part of the standardization programme is the establishment of a central data bank, to which all users of the STRES battery are encouraged to contribute as a matter of routine.

The major functions of data exchange will be (a) to help to identify the psychometric properties of the tests, (b) to provide normative data, (c) to indicate the pattern of performance change associated with a particular stressor, (d) to indicate the effects of a range of stressors on a particular mental process, (e) to indicate the effects of "incidental" variables such as age on mental performance, (f) to reveal occupational differences in performance that may be relevant to selection issues, and (g) to facilitate communication between users with common interests.

The location of the databank, and the means of gaining access to it, have yet to be decided. This information will be disseminated as soon as it is available. In the meantime, the databank will be run by the members of AGARD Working Group 12 who volunteered to do this. Searchers wishing to enter new data or to interrogate the database should contact the nearest Working Group member volunteer (see Appendix A).

Data should be transmitted by means of floppy disks. Files should be provided in ASCII code using the Data Interchange Format (DIF). Data should be stored on double-sided 5.25 inch MS-DOS diskettes (40 tracks, 9 sectors) or 3.5 inch MS-DOS diskettes (80 tracks, 9 or 18 sectors); these storage media were selected because they are available to nearly all laboratories.

Part 1 of the file contains general information, and Part 2 performance data. Information to be entered as Part 1 includes author(s), keywords, reference if published, date of the experiment, aims, design, subjects, and deviations from standard conditions if any.

The software associated with the STRES battery will include (a) a routine to collect and store, prior to administration of the battery, the general information comprising Part I of the data transfer file, and (b) a routine to set up the data transfer files in the required format. Further details are published in AGARDograph 308.

## Conclusions and Prospect

The objective of the STRES Battery is to provide a solid core of well-accepted performance tests for use by the applied researcher. As more data become available, the battery may prove useful as a yardstick or an anchor point to those who wish to develop new approaches. Further, the accumulation of data will permit the validity of the STRES battery to be more fully explored, in addition to the formal validation studies as outlined in the section on psychometrics. Data accumulation will also permit an examination of the range of stressors to which a given task is sensitive. It will also be possible to investigate the extent to which each test is sensitive to individual differences, and the relevance of the test to occupational groups.

Further progress is dependent upon acceptance of the battery by research establishments, and data exchange between laboratories. Cooperation and enthusiasm of test users is very important. In this respect, the favourable responses to the surveys conducted by Sanders et al. (1986) and by AGARD AMP Working Group 12 are encouraging.

## References

1 Sanders, A. F., Haygood, R. C., Schroiff, H.-W., & Wauschkuhn, C. H. Standardization of Performance Tests: A Proposal for Further Steps. Aachen: Institute for Psychology, RWTH, Scientific Report Phase I (1 Sept 1985-31 Aug 1986) prepared for USAF Air Force Office of Scientific Research and European Office of Aerospace Research and Development, 1986.

2 Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

3 Sternberg, S. Two operations in character recognition: Some evidence from reaction time measurements. Perception and Psychophysics, 1967, 2, 45-53.

4 Hockey, R., & Hamilton, P. The cognitive patterning of stress states. In R. Hockey (Ed.), Stress and Fatigue in Human Performance. Chichester: John Wiley & Sons, 1983.

5 Gaillard, A. W. K., & Steyvers, F. J. J. M. (in press). Sleep loss and sustained performance. In A. Coblentz (Ed.), Vigilance and performance in automatized systems. Dordrecht (Netherlands): Nijhoff.

6 Sanders, A.F. Towards a model of stress and human performance. Acta Psychologica, 1983, 53, 61-97.

7 Boer, L.C., Gaillard, A.W.K., & Jorna, P.G.A.M. [Taskomat, A task battery for information processing.] In Dutch. Soesterberg, Netherlands: TNO Institute for Perception, Report no. IZF 1982-2.

8 Shingledecker, C. A. A Task Battery for Appli· ı Human Performance Assessment Research. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratories Report No. AFAMRL-TR-84-071, 1984.

9 Englund, C. E., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., Wilson, K. P., & Hegge, F. W. Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB). 1: Design and Specification of the Battery. San Diego, California: Naval Health Research Center Report no. 87-10, 1987.

10 Donders, F.C. [On the speed of mental processes.] In W.G. Koster (Ed. and trans.), Attention and performance II, 412-431. Amsterdam: North Holland, 1969. (Reprinted from Acta Psychologica, 1969, 30.)

11 Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and performance II, 276-315. Amsterdam: North Holland, 1969. (Reprinted from Acta Psychologica, 1969, 30.)

12 Chiles, W. D., & Alluisi, E. A. On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 1979, 21, 515-528.

13 Chiles, W. D., Bruni, C. B., & Lewis, R. A. Methodology in the Assessment of Complex Human Performance: The Effects of Signal Rate on Monitoring a Dynamic Process. Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration Report no. FAA-AM-69-9, 1969.

14 Chiles, W. D., & Jennings, A. E. Effects of alcohol on complex performance. Human Factors, 1970, 12, 605-612.

15 Hall, T. J., Passey, G. E., & Meighan, T. W. Performance of Vigiliance and Monitoring Tasks as a Function of Workload. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratories Report No. AMRL-TR- 65-22, 1965.

16 Chiles, W. D., Jennings, A. E., & Alluisi, E. A. The Measurement and Scaling of Workload in Complex Performance. Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration Report no. FAA-AM-7834, 1978.

17 Ashcraft, M. H., & Stazyk, E. H. Mental addition: A test of three verification models. Memory and Cognition, 1981, 9, 185-196.

18 Hitch, G. J. The role of short term working memory in mental arithmetic, Cognitive Psychology, 1978, 10, 302-323.

19 Ross, J. Extended practice with a single-character classification task. Perception and Psychophysics, 1970, 8, 276-278.

20 Kristofferson, M. W. When item recognition and visual search functions are similar. Perception and Psychophysics, 1972, 12, 379-384. (a)

21 Graboi, D. Searching for targets: The effects of specific practice. Perception and Psychophysics, 1971, 10: 300-304.

22 Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 1977, 84, 1-66.

23 Shiffrin R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 1977, 84, 127-190.

24 Kristofferson, M. W. Effects of practice on character classification performance. Canadian Journal of Psychology, 1972, 26, 540-560. (b)

25 Lohman, D. F. Spatial Ability: A review and Reanalysis of the Correlational Literature. Stanford, California: Aptitude Research Project, School of Education, Stanford University, Technical Report no. 8, 1979.

26 Fitts, P. M., Weinstein, M., Rappaport, M, Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern perception: A probability approach. Journal of Experimental Psychology, 1956, 51,1-11.

27 Chiles, W. D., Alluisi, E. A., & Adams, O. S. Work schedules and performance during confinement. Human Factors, 1968, 10, 143-196.

28 Wickens, C. D. Engineering Psychology and Human Performance. Columbus, Ohio: Charles E. Merrill Publishing Co., 1984.

29 Jex, H. R., McDonnell, J. D., & Phatak, A. V. A "critical" tracking task for manual control research. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, 138-144.

30 Ashkenas, I. L., & McRuer, D. T. The Determination of Material Handling Quality Requirements from Airframe/Human-Pilot System Studies. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center Report no. TR 59-135, 1959.

31 Jex, H. R., & Cromwell, C. H. Theoretical and Experimental Investigation of Some New Longitudinal Handling Quality Parameters. Wright-Patterson Air Force Base, Ohio: Aeronautical Systems Division Report no. TR 61-26, 1961.

32 Baddeley, A. D. A 3-minute reasoning test based on grammatical transformations. Psychonomic Science, 1968, 10, 341-342.

33 Slobin, D. I. Grammatical transformations and sentence comprehension in childhood and adulthood. Journal of Verbal Learning and Verbal Behaviour, 1966, 5, 219-227.

34 Wason, P. C. Response to affirmative and negative binary statements. British Journal of Psychology, 1961, 52, 133-142.

35 Baddeley, A. D., & Hitch, G. J. Working memory. In G. H. Bower (Ed.), The Psychology of Learning and Motivation Volume 8. New York: Academic Press, 1974.

36 Hitch, G. J., & Baddeley, A. D. Verbal reasoning and working memory. Quarterly Journal of Experimental Psychology, 1976, 28, 603-621.

37 Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. Evidence for a visuo-spatial scratch-pad in working memory. Quarterly Journal of Experimental Psychology, 1986, 38A, 675-688.

38 Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

39 Shingledecker, C. A., Acton, W. H., & Crabtree, M. S. Development and application of a criterion task set for workload metric evaluation. Paper presented at the Second Aerospace Behavioral Engineering Technology Conference, (No. 83/419), Long Beach, California, 1983.

40 Michon, J. A. Tapping regularity as a measure of perceptual motor load. Ergonomics, 1966, 9, 401-412.

41 Vidulich, M., & Wickens, C. D. Time-sharing, manual control and memory search: The joint effects of input and output modality competition, priorities and control order. Champaign (Ill): Department of Psychology, University of Illinois, Report no. EPL-81-4, 1981.

42 Wickens, C. D., & Sandry, D. Task-hemispheric integrity in dual task performance. Acta Psychologica, 1982, 52, 227-247.

43 Boer, L. C., Ruzius, M. H. B., Mimpen, A. M., Bles, W., & Janssen, W. H. Psychological Fitness During a Manoeuvre. Soesterberg: TNO Institute for Perception Report no. IZF 1984-17, 1984.

## Appendix A

Members of AGARD AMP Working Group 12
volunteering for initial database management

Col G. Santucci (Chairman)
C.E.R.M.A.
26 Boulevard Victor
75996 Paris
France
Tel.: (33)-1-4552-6014
Telex: AIR CERMA PARIS 200 795

Dr L.C. Boer
TNO Institute for Perception
Kampweg 5
3769 ZG Soesterberg
The Netherlands
Tel.: (31)-3463-56463
Telefax: (31)-3463-3977

Dr E.W. Farmer
RAF Institute of Aviation
Medicine
Farnborough
Hants GU14 6SZ
UK
Tel.: (44)-252-24461 extension 4366

Dr K.-M. Goeters
DFVLR
Department of Aviation and
Space Technology
Sportallee 54
2000 Hamburg 63
West Germany Tel.: (49)-40-513096-0

Dr J.D. Grisset
Naval Aerospace Medical
Research Laboratory
Pensacola
Florida 32508
USA
Tel.: (1)-904-452-4457

Mr E.O. Schwartz
DFVLR
Institute for Aerospace Medicine
Postfach 90 60 58
5000 Cologne 90
West Germany
Tel.: (49)-2203-601-3169 or 3076
Telefax: (49)-2203-67310
Telex: 8874410 (dfvw d)

Dr A. Wetherell
Chemical Defence Establishment
Porton Down
Salisbury
Wiltshire SP4 0JQ
UK
Tel.: (44)-980-610211 extension 411

Dr G.F. Wilson
AAMRL/HEG
Wright-Patterson AFB
Ohio 45433
USA
Tel.: (1)-513-255-8748 or 7586
Telefax: (1)-513-255-9198

# DISCUSSIONS OF SESSION IV

BILLINGS: There has been a rather curious set of contradictions here today. We have heard that personality traits are fixed and that is part of the definition of them, yet we have all experienced over many years a situation in which highly competent military pilots, with certain supposedly fixed personality traits, have formed later in their lives the major source of supply for our civil air transport pilot community. The best and most effective members of which have quite a different set of presumed personality traits for success. I am having a little difficulty with that. I have a considerable difficulty with it with respect to Scandinavia. The vast majority of whose military pilots are fighter pilots. Virtually all of whose air transport pilots have, at least until very recently, come from that same largely fighter pilot community.

I think we have a conceptual problem here, and I think we need to start working on it. Perhaps Dr Kemmler gave us a hint of it when he emphasised the need for training as opposed to selection and pointed out that what we select may be quite a different "cat" in five, ten or fifteen years' time. I think we have seen here today some of the ways in which training may compensate both in part for our inability to select on certain traits, perhaps our lack of knowledge of what traits to select on, perhaps our lack of knowledge as to whether they are, in fact, traits at all. We have heard about dual task performance as a predictor, and apparently a rather successful one, in The Netherlands of success in handling the Orion in its military roles. The State of Israel's Air Force has recently done a very interesting study on computer games as a prelude to flight training, and the results were fairly dramatic. I think it a tiny step between using them as a prelude to flight training and using them as a predictor of the likelihood of success in flight training.

Perhaps models like those that were discussed by Dr Damos this morning are, in fact, coming close enough to maturity to begin to govern some of our selection and more importantly, I think, a good bit of what we do in training. It is quite clear that our training is fairly successful. It is also quite clear that we do not know what we are doing in many respects. We are doing what has worked in the past. We are doing it almost entirely empirically, and I think we are selling ourselves short. We know more than that now. I think it has been an excellent conference by virtue of reminding us of that and perhaps by reminding us of some of the contradictions that continue to persist in our world of thinking. We do need to step back from some of these trees and look rather more carefully at the woods when we are describing the forest.

URSIN: I think it is very important, if you are going to use personality traits in prediction, that you are very accurate in stating what you are going to predict. Likewise, when you are selecting, that you are very accurate in what you want to select. This afternoon we have had some controversy regarding the Defence Mechanism Test from Scandinavia, but it is quite clear, as far as I can see, that there is no controversy at all, at least not from the Norweigan point of view. We published in 1978 the first paper from my own group on the DMT. Our principle finding was that the DMT does not predict the results of training, that is, if you are talking about performance in the ordinary way that performance is scored by officers watching what the pilot is doing. There is no correlation at all. On the other hand, there is no correlation on that aspect of performance and what happens when the trainee is faced with life threatening situations. Our main point has always been that the DMT is there to be used if you want to predict what will happen when an individual's life is threatened. That is very different from an ordinary test situation in which performance is assessed. We made the point again in Vaernes's paper in 1982 and we have repeatedly made the same point in our the papers over the last 3 or 4 years.

As far as divers, trainee parachutists and pilots are concerned, the DMT does not predict anything in the training situation. It does, however, predict errors made in situations where the person is extremely activated. This morning Vaernes showed the type of data, the ordinary performance, that the DMT does not predict, but it does predict performance in the simulator when the man is really put under pressure. So if we are selecting personnel for the everyday routine, stay away from the DMT. If, on the other hand, we are selecting for what could happen if there were to be an accident then you should use the DMT. If we are selecting pilots for their life time performance, well then, we are not going to look only at their performance as fighter pilots. Most of these people, as they have done in the last 20-30 years, are going to have a long career manoeuvring aeroplanes when they almost never, or perhaps never, come up against situations where they require the type of attributes which are probed by the DMT.

There is one exception to this. If we are talking about the life quality of a crew or of a group of people collaborating, then it seems as if the DMT is picking up something. Then again you cannot do it in an acute test situation. Remember Foushee's honeymoon effect. It is what happens after years of collaboration that things seem to pick up. In the last 2 or 3 issues of 'Work and Stress' there are again papers on DMT that seem to point out that the Defence Mechanisms, as assessed by this test, actually have something to do with how we live together. So if that is what we are going to select for, you should select a test for that. We cannot agree and we cannot do an adequate job unless we are very sure we know what kind of traits we are going to test for, what kind of performance is relevant and in what kind of situation, and whether these decisions that we are going to make are hot or cold.

STEVENSON: With respect to the remarks just made when we were talking about test saturation, there are two distinct environments. One is an airline pilot who has a multiple task environment that he has to deal with on a day to day basis. The other is a military pilot who also has a multiple task environment that is increasing everyday even in peacetime. There is also, and this in my experience is what is killing our pilots in Europe, a multiple task situation in which an emergency, a life threatening situation, interfers with their normal ability to process situations. What I have heard many times in the conference has been about experiments done in simulators and on computer keyboards where there was absolutely no stress in terms of a life threatening event or even a threat to the person's personality. Although I am a little naive in psychological methods, is there any way that you scientists can introduce risk into an experimental situation which might shed some light on peoples' abilities to deal with these life threatening situations in real life? What I am suggeting, for instance, is that you offer a naive subject at a computer keyboard a financial reward and take away a certain amount of money every time he makes an error. So that there is a monetary penalty if he does not perform well. For a military aviator, if you are going to do a study on him in a simulator, do it during a graded simulator ride where his superiors are observing him and his professional standing may have something to do with his performance. I think such a technique might gain a little more insight into how truly threatening situations are dealt with.

# FINAL COMMENTS

**ANDERSEN**

During this symposium we have talked extensively of flying in general, but somewhat less about military aerospace operations. Returning to my opening statement on Tuesday morning about aggressiveness, I still believe that "controlled aggressiveness" is indispensable for effective use of military air power.
The Air Force task is not just to fly. The purpose of military flying is to fight, or to support other fighters to win their battles. There exist no other justification for an Air Force to be trained and supported. If not seriously aiming for winning fights Air Forces become reduced to expensive toys.
This makes us revisit the philosophy of selection and training. Do we select and train for a good record of flying safety in peace time or do we select and train in order to create a fighting force for wartime purposes? A good safety record is nice, but it must not become a goal of its own even during peacetime exercises. If we loose the main objective from our list of priorities the entire Air Force may be lost within 24 hours of engagement in a major conflict. Military air power usage is a question of formulating winning strategies. It is, therefore, a question of what to be prepared for and, moreover, what not to be surprised by. The very big surprises, as we have heard, are what cause accidents in aerospace operations.

This finally brings me to the importance of scientific methods when we select, train and treat air crew. Science is about defining problems which may be successfully worked upon by discriminating methods. The results have to be reliable and reproducible and the scientist has to be able to explain what they mean. Possibly, at least in the behavioural sciences, questions never changes, only the answers'.

I thank you all very much for your participation. It has been interesting for those of us who have been planning the programme. I hope you have enjoyed it.

C-2

# REPORT DOCUMENTATION PAGE

| 1. Recipient's Reference | 2. Originator's Reference | 3. Further Reference | 4. Security Classification of Document |
|---|---|---|---|
| | AGARD-CP-458 | ISBN 92-835-0517-4 | UNCLASSIFIED |

**5. Originator** Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

**6. Title** HUMAN BEHAVIOUR IN HIGH STRESS SITUATIONS IN AEROSPACE OPERATIONS

**7. Presented at** the Aerospace Medical Panel Symposium held in The Hague, The Netherlands from 24—28 October 1988

| 8. Author(s)/Editor(s) | 9. Date |
|---|---|
| Various | June 1989 |

| 10. Author's/Editor's Address | 11. Pages |
|---|---|
| Various | 212 |

**12. Distribution Statement** This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

**13. Keywords/Descriptors**

| | | |
|---|---|---|
| Aircrew | Response | Emergency |
| Stress | Behaviour | Performance |
| Accident | Personality | Risk |
| Psychological assessment | Selection | |


**14. Abstract**

These Proceedings include the Technical Evaluation Report, the Keynote Address, 20 papers and ensuing discussions from the Symposium sponsored by the AGARD Aerospace Medical Panel held in The Hague, The Netherlands from 24—28 October 1989.

As the human operator is more and more clearly shown to be the limiting factor in the operational performance of modern aerospace systems, more and more rigorous selection criteria must be applied for psychological as well as medical selection of aircrew.

This Symposium examined this problem from the point of view of human behaviour in high stress situations, specifically looking at incident and accident experience, personality traits, responses to stress and prediction of behavioural responses.

These proceedings will be of interest to those involved with the psychological selection and/or assessment of aircrew.

AGARD Conference Proceedings No.458
Advisory Group for Aerospace Research and Development, NATO
HUMAN BEHAVIOUR IN HIGH STRESS SITUATIONS IN AEROSPACE OPERATIONS
Published June 1989
212 pages

These Proceedings include the Technical Evaluation Report, the Keynote Address, 20 papers and ensuing discussions from the Symposium sponsored by the AGARD Aerospace Medical Panel held in The Hague, The Netherlands from 24—28 October 1989.

As the human operator is more and more clearly shown to be the limiting factor in the operational performance of

P.T.O

AGARD-CP-458

Aircrew
Stress
Accident
Psychological assessment
Response
Behaviour
Personality
Emergency
Performance
Risk
Selection

AGARD Conference Proceedings No.458
Advisory Group for Aerospace Research and Development, NATO
HUMAN BEHAVIOUR IN HIGH STRESS SITUATIONS IN AEROSPACE OPERATIONS
Published June 1989
212 pages

These Proceedings include the Technical Evaluation Report, the Keynote Address, 20 papers and ensuing discussions from the Symposium sponsored by the AGARD Aerospace Medical Panel held in The Hague, The Netherlands from 24—28 October 1989.

As the human operator is more and more clearly shown to be the limiting factor in the operational performance of

P.T.O

AGARD-CP-458

Aircrew
Stress
Accident
Psychological assessment
Response
Behaviour
Personality
Emergency
Performance
Risk
Selection

AGARD Conference Proceedings No.458
Advisory Group for Aerospace Research and Development, NATO
HUMAN BEHAVIOUR IN HIGH STRESS SITUATIONS IN AEROSPACE OPERATIONS
Published June 1989
212 pages

These Proceedings include the Technical Evaluation Report, the Keynote Address, 20 papers and ensuing discussions from the Symposium sponsored by the AGARD Aerospace Medical Panel held in The Hague, The Netherlands from 24—28 October 1989.

As the human operator is more and more clearly shown to be the limiting factor in the operational performance of

P.T.O

AGARD-CP-458

Aircrew
Stress
Accident
Psychological assessment
Response
Behaviour
Personality
Emergency
Performance
Risk
Selection

AGARD Conference Proceedings No.458
Advisory Group for Aerospace Research and Development, NATO
HUMAN BEHAVIOUR IN HIGH STRESS SITUATIONS IN AEROSPACE OPERATIONS
Published June 1989
212 pages

These Proceedings include the Technical Evaluation Report, the Keynote Address, 20 papers and ensuing discussions from the Symposium sponsored by the AGARD Aerospace Medical Panel held in The Hague, The Netherlands from 24—28 October 1989.

As the human operator is more and more clearly shown to be the limiting factor in the operational performance of

P.T.O

AGARD-CP-458

Aircrew
Stress
Accident
Psychological assessment
Response
Behaviour
Personality
Emergency
Performance
Risk
Selection

modern aerospace systems, more and more rigorous selection criteria must be applied for psychological as well as medical selection of aircrew.

This symposium examined this problem from the point of view of human behaviour in high stress situations, specifically looking at incident and accident experience, personality traits, responses to stress and prediction of behavioural responses.

These proceedings will be of interest to those involved with the psychological selection and/or assessment of aircrew.

ISBN 92-835-0517-4

modern aerospace systems, more and more rigorous selection criteria must be applied for psychological as well as medical selection of aircrew.

This symposium examined this problem from the point of view of human behaviour in high stress situations, specifically looking at incident and accident experience, personality traits, responses to stress and prediction of behavioural responses.

These proceedings will be of interest to those involved with the psychological selection and/or assessment of aircrew.

ISBN 92-835-0517-4

modern aerospace systems, more and more rigorous selection criteria must be applied for psychological as well as medical selection of aircrew.

This symposium examined this problem from the point of view of human behaviour in high stress situations, specifically looking at incident and accident experience, personality traits, responses to stress and prediction of behavioural responses.

These proceedings will be of interest to those involved with the psychological selection and/or assessment of aircrew.

ISBN 92-835-0517-4

modern aerospace systems, more and more rigorous selection criteria must be applied for psychological as well as medical selection of aircrew.

This symposium examined this problem from the point of view of human behaviour in high stress situations, specifically looking at incident and accident experience, personality traits, responses to stress and prediction of behavioural responses.

These proceedings will be of interest to those involved with the psychological selection and/or assessment of aircrew.

ISBN 92-835-0517-4

AGARD

NATO OTAN

7 rue Ancelle · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone (1)47.38.57.00 · Telex 610 176

## DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres.Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

### NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defen
Ved Idraetspa
2100 Copenh

FRANCE
O.N.E.R.A. (I
29 Avenue de
92320 Châtill

GERMANY
Fachinformati
Physik, Mathe
Karlsruhe
D-7514 Eggen

GREECE
Hellenic Air F
Aircraft Suppo
Department of
Holargos, Athe

ICELAND
Director of Avi
c/o Flugrad
Reyjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3 Piazzale Adenauer
00144 Roma/EUR

LUXEMBOURG
*See* Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
2600 AC Delft

NORWAY

Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Research Center
M/S 180
Hampton, Virginia 23665

NASA
National Aeronautics and
Space Administration
Washington, D.C.
20546

SPECIAL FOURTH CLASS MAIL
BOOK

Postage and Fees Paid
National Aeronautics and
Space Administration
NASA-451

Official Business
Penalty for Private Use $300

L2 001 AGARDCP4588909155002672D
DEPT OF DEFENSE
DEFENSE TECHNICAL INFORMATION CENTER
DTIC-FDAC
CAMERON STATION BLDG 5
ALEXANDRIA VA 223046145

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

### PURCHASE AGENCIES

National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Branch
NASA Headquarters (NIT-40)
Washington D.C. 20546, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA

SPS

*Printed by Specialised Printing Services Limited*
*40 Chigwell Lane, Loughton, Essex IG10 3TZ*

ISBN 92-835-0517-4