



## A REVIEW OF IMPORTANT COGNITIVE CONCEPTS IN AVIATION

Ana P. G. MARTINS

*Department of Pilot Assistance, Institute of Flight Guidance, German Aerospace Centre (DLR),  
Lilienthalplatz 7, 38108 Braunschweig, Germany  
E-mail: ana.martins@dlr.de*

*Received 10 May 2016; accepted 8 June 2016*

**Ana P. G. MARTINS, Dr**

*Education: Purdue University, 2006 MSc in Learning and Memory. Purdue University, 2009 PhD in Learning and Memory.*

*Affiliations and functions: since 2013 – German Aerospace Centre DLR, researcher.*

*Research interests: stress and workload reduction in the cockpit, acceptance of new aircraft technologies.*

*Publications: author of 3 papers; coauthor of 4 papers. Coauthor of 3 book chapters.*

**Abstract.** Even considering the current low accident rate in aviation, the anticipated growth in the number of airplanes in the air in the next decades will lead to an inadmissible rise in the number of accidents. These have been mostly attributed to human error and a misunderstanding of automation by the crew, especially during periods of high workload and stress in the cockpit. Therefore, increased safety requires not only advances in technology, but improved cockpit design including better human-machine interface. These cannot be achieved however, without considering some of the cognitive constructs that affect the behaviour of pilots in the cockpit. In fact, given its characteristics and public visibility, the flight deck of commercial jets is one of the most common arenas for the study of complex and skilled human performance. Here I present a literature review on the selected topics of workload, situation awareness, stress and automation in the cockpit, with the goal of supporting the development of new technologies.

**Keywords:** workload; situation awareness; stress; automation, pilot, cockpit.

### 1. Introduction

Considerable efforts by the aviation community have resulted in the achievement of a safety record in air transport that is unmatched by other modes of transport. Yet, for the 10-year period between 2005 and 2014 there were around 400 accidents with substantial aircraft damage, hull loss, serious injury or death. Boeing statistics further indicate that in the year 2014 alone there were over 25.6 million departures, with one accident occurring, on average, every 12 days worldwide (Boeing 2014). Over the next 20 years air travel is expected to grow around 5 percent per year, resulting in an estimated increase in the number of airplanes in service in 2034 to around 43500 (Boeing 2015) or 38500 (Airbus 2015). Assuming accident rate remains the same, overall frequency of accidents and number of fatalities, while still relatively low, will translate into several major incidents and accidents per week, an unacceptable trend to the general public.

With current aircraft technology, present safety levels can only be assured through mature crew resource management that focus on the concept of pilots sharing information, and complementing and cross-checking each other. As such, regulations require that commercial airlines have at least two pilots aboard an airplane, including two in the cockpit. In today's operational environment typically one pilot flies the aircraft (pilot-flying), while the other (pilot-not-flying) manages systems and provides support, such as radio communications and trouble-shooting in the event of failures. The pilot-not-flying is expected not only to monitor the aircraft, but also the actions of the pilot-flying, pointing out if something is not right, or suggest that some tasks be offloaded from the pilot-flying (Vidulich 2002). Thus, even though a single pilot is technically capable of safely flying a large aircraft from take-off to landing, this cannot be done within the expected safety margins in

today's operational context, let alone those foreseen in the medium-term future.

Whenever unexpected events occur that require two or more tasks to be performed concurrently, it might lead to stress, high workload levels and a loss of situation awareness. In those cases, the pilot must prioritize and decide which tasks to perform first (Morris, Leung 2006). According to Loukopoulos *et al.* (2003), airline formal procedures and training do not provide clear guidance on how to manage concurrent demands. For example, it is expected that during taxi the pilot-not-flying not only monitors the actions of the other pilot, but also responds to radio calls and finishes up any performance calculations left-over from the preflight phase. Interruptions to ongoing tasks from radio communications and the cabin crew are also quite common, and pilots must remember to return to the interrupted task later. Deferred actions, however, are particularly vulnerable to forgetting (Dismukes, Nowinski 2006).

As aircraft equipment becomes progressively more reliable, the percentage of accidents attributed to human error is expected to increase. According to Darby (2006), 55% of 134 major crashes that occurred between 1996–2005 could be directly linked to human error. Although mistakes and inappropriate reactions to events can be made in all phases of flight, they are more likely to occur when the workload in the cockpit is high, i.e. when the cockpit crew is required to perform very demanding or a large number of actions in a limited period of time. Between 2005 and 2014 approximately half of all fatal accidents (48%) occurred during the approach and landing phases, with another 13% taking place during take-off and initial climb. These correspond to the phases when the crew is dealing with the highest workload and the airplane is in closer proximity to the ground, reducing manoeuvre margins in case of an emergency (Boeing 2014). As such, the reduction in the overall number of accidents may require advances in technology, including higher levels of automation and improved cockpit displays to help the cockpit crew manage peak workload situations. This would ensure that pilots have the opportunity to address all relevant issues in a timely and appropriate manner.

## 2. Workload

Studies in workload started as early as 1930, but evaluated tasks mostly had a physical component that required the manipulation of machines (Sheridan, Simpson 1979). With increased automation in workplaces as a result of the introduction of computers, it became harder to define workload, especially when its demands are transient: a monitoring situation may, suddenly, require the processing of several error messages before a single response can be made, after which the

operator reverts back to monitoring. Even though there is no universally accepted definition of mental workload (Cain 2007), in its simpler form it refers to the measurement of the mental processing demands placed on a person during the performance of a task (Gopher, Donchin 1986). In other words, mental resources or capacities are needed in order to complete an assignment, and those available to the human operator may not always match those required to satisfactorily perform the task (Gopher, Braune 1984).

Mental resources or capacities refer to the human information processing system. Stimuli arriving at the sensory receptors need to be perceived and translated to a response. The classic psychological model identifies three main stages (Fig. 1): perception, decision making/response selection and response programming/execution (Johnson, Proctor 2004). Attentional and memory systems are assumed to affect these stages and each will be briefly discussed later on.

One of the earliest theories of mental workload was developed by Kahneman in 1973. His was a *unitary-resource model of attention*, which viewed resources as an undifferentiated, limited pool, to be shared by concurrent tasks (Wickens 2008). According to this model, whenever demand for attentional resources exceeds availability, performance should suffer. Nevertheless, several experimental findings, such as the difficulty insensitivity effect, the structural alteration effect and the perfect-time sharing effect, suggested that under some conditions two tasks appear to demand separate resources (Johnson, Proctor 2004). The *difficulty insensitivity effect* refers to the finding that performance is not always degraded on one task when difficulty levels increase on the concurrent one (Kantowitz, Knight 1976). The *structural alteration effect* corresponds to differences in performance on one task depending on the response format on the other (for example, McLeod (1977) reported that performance in a tracking task was affected by whether the second task required a manual or verbal response). Finally, the *perfect-time sharing effect* occurs when two tasks that interfere with other tasks do not affect each other when performed together (e.g., Schumacher *et al.* 2001). None of these effects can be explained by a unitary-resource model, which predicts that any task should always demand the same resources and, consequently, that two tasks would consistently interfere with each other in the same way. These and other findings have led to the development of the concept of multiple resources.

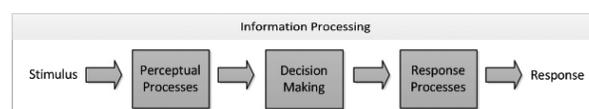


Fig. 1. The classic information processing framework

Developed by Wickens (1980), the *Multiple Resources Theory* is one of the most influential theories to address these particular issues and explain performance in high workload dual-task situations. Its success lies on its neurophysiological validity (there is some parallel between the separate dimensions of human information processing and specific brain structures), at the same time it provides some guidelines to human factors designers regarding the configuration of a task or system (Wickens 2008).

Wickens (1980) identified four dimensions, each with two discrete levels. The first dimension corresponds to two information processing *stages*: resources involved with perceptual-cognitive activity are functionally different from those related to response processes (Wickens 1991). For instance, a task that requires display reading or voice comprehension (both perceptual-cognitive activities) can be time-shared efficiently with another task that involves activating a switch or pressing a button (both being response processes) (Wickens, Liu 1988). Or, as seen in commercial aircraft, a pilot is perfectly able to listen to instructions from the air traffic control (ATC) and, at the same time, tune and identify a frequency of the Instrument Landing System. The second dimension refers to the resources used in processing spatial and verbal *codes*, which are not shared and are in fact associated with the different cerebral hemispheres. This dichotomy is seen at the levels of perception (graphics vs. speech), cognition (spatial working memory vs. memory for linguistic information) and response (manual responses vs. speech). An example of the later is the reduced interference between a task that uses a control stick (manual responses usually involve spatial codes) and another that requires giving a response to ATC on the radio (verbal code).

The third dimension consists of the different perceptual (visual or auditory) *modalities*. It is a well-known fact that it is easier to attend to both visual and auditory sources than to two simultaneous auditory or visual messages. Some studies, however, have called into question the strength of the modality dimension (Wickens 1991). Wickens and Liu (1988) found that performance on a continuous visual task suffered more in the presence of a discrete auditory task than if the concurrent task was presented visually. The authors called it a pre-emption effect, because the auditory stimulus shifted attention (thus, resources) away from the continuous visual task. Latorella (1999) showed in a simulated flight deck that, during visual tasks, auditory interruptions were more disruptive than visual interruptions. Also, pilots are often seen to interrupt their actions to answer the air traffic controllers, a finding confirmed by Morris and Leung (2006). This may lead to errors, such as skipping a checklist item when resuming the list. Humans are highly predisposed to attend to auditory stimuli given the limited

capacity ( $7 \pm 2$  digits) of short-term memory (the type of memory retained for only a very brief period). As such, when presented with auditory stimuli which will be easily and quickly forgotten, they tend to pay attention to them, at the expense of other concurrently presented visual stimuli. In fact, Huey and Wickens (1993) reported that tasks that require short-term memory are often perceived as imposing greater workload than tasks that require the retrieval of information from long-term memory. Finally, another problem with the modality dimension was reported by Gladstones *et al.* (1989), who showed that when two discrete tasks were being performed at their maximum speed, subjects' capacity to process information was unaffected by the perceptual modality used (same or different).

The most recent addition to the model, the fourth dimension, corresponds to the focal and ambient visual *channels*. Focal vision is required for pattern/object recognition and high acuity perception, whereas ambient vision is involved in orientation and movement perception of oneself. An example of using both resources is reading a map while flying in visual meteorological conditions, that is keeping an eye on the outside world looking for traffic while at the same time focusing attention on the map.

Despite the criticisms (e.g., Wickens 1991, 2008), the Multiple Resources model has received some support from experiments that examined dual-task performance employing new cockpit technology for the display of traffic information and flight parameters (Wickens *et al.* 2003; Wickens, Colcombe 2007).

Several methods developed to measure workload are essentially measures of arousal. Arousal can be generally defined as a physiological sense of readiness to act. The relationship between performance and arousal levels was first described by Yerkes and Dodson (1908), in what is now called the *Yerkes-Dodson law*. It states that a high level of arousal can enhance performance on an easy task, but on a difficult task performance is an inverted U-shaped function of arousal (Fig. 2). Additionally,

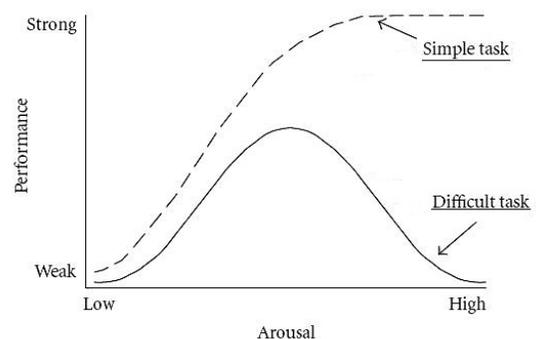


Fig. 2. Graphic representation of the Yerkes-Dodson law (adapted from Diamond *et al.* (2007)).

the more difficult the task is, the lower the arousal level at which performance peaks. According to the *cue-utilization theory* (Easterbrook 1959), high arousal leads to a decrease in the number of monitored cues, and is thus beneficial only when few cues are presented.

In situations of high workload, performance levels can sometimes be maintained at the expense of extra mental effort, which may lead to feelings of fatigue, stress or strain. One way to manage high workload is to make a *strategy adjustment* or *strategy shift*. That is, the operator might change his or her behaviour to a less effortful strategy to perform the task. In general, the chosen strategy depends on three factors: the individual characteristics (training, motivation, age, health, etc.), task characteristics (i.e., its requirements, including work conditions) and workload levels (Sperandio 1971).

A change in strategy by itself might indicate a change in workload. Sperandio (1971) described a study done with air traffic controllers which showed that as workload increased, their operative methods changed. More specifically, with low workload levels the controllers had each airplane follow the shortest way between the ATC sector entrance-point and the runway threshold. As workload increased, they adopted a less efficient, more time consuming strategy of sending the aircraft to a standard route (which included the use of holding patterns followed by standard procedures to land). The phenomenon of *attentional narrowing* is another example of a strategy shift. It refers to the finding that under a stressful situation, humans sometimes restrict their attention to some specific cues or information sources (those considered more important), ignoring others. This is an efficient approach to reduce workload in the short term, but it might lead to errors and omissions. An often cited example of this effect occurred with the Eastern Airlines flight 401, which crashed because both pilots were focusing on a malfunction of the landing gear light instead of flying the airplane. Even only telling people that they will experience a stressful situation, without actually going through it, is enough to cause a significant reduction in the number of attended stimuli (Weltman *et al.* 1971). Another compensatory mechanism is the *fatigue after-effect*. In this case, after a particularly stressful task the operator may switch to low cost strategies in subsequent tasks (Schellekens *et al.* 2000). As an example, Stewart *et al.* (2006) reported a study where the performance error rate of pilots from a commercial airline decreased towards the end of a week of rosters. The reason was simple: as the pilots got tired, they compensated by using more automation. High mental workload demands can also produce *speed-accuracy trade-offs*. Reaction time and accuracy are known to be inversely related, that is as speed increases there usually is a decrease in task accuracy. The trade-off is said to occur when fast reaction

times are accompanied by high error rates. This corresponds to a risky strategy by the operator, as opposed to a more conservative approach that leads to fewer errors, but also an increase in response times (Sperandio 1971).

Low workload is almost as crucial as high workload levels. This is often associated with fewer responses, slower response times, poor decision making, loss of situation awareness, change in decision criteria and a failure to detect relevant signals, mostly due to boredom, distraction and reduction in vigilance or attention (e.g., Hancock, Warm 1989; Huey, Wickens 1993; Parasuraman, Mouloua 1987). Mackworth (1948), for example, reported that vigilance tasks that last longer than 30 minutes lead to a significant decline in reaction times and accuracy levels.

As Casali and Wierwille (1983) noted, mental workload must be inferred as it cannot be directly observed. Thus, three broad categories of measures have been developed: physiological, performance-based and subjective. According to Meister and Gawron (2010), *physiological measures* in aviation are mostly used in experiments that study the effects of acceleration, hypoxia, noise level, fatigue, alcohol, drugs, or workload. High workload levels are usually associated with cardiovascular and respiratory changes, as well as with specific patterns in brain electrical activity. Changes in heart rate within brief time intervals (measured as heart-rate variability), as well as changes in pupil diameter, are two of the most common body responses measured – high workload leads to large pupil size and a decrease in heart-rate variability (Beatty 1982; Van Orden *et al.* 2001; Van Roon *et al.* 2004; Vicente *et al.* 1987). Moreover, event-related potentials computed by averaging the electroencephalography (EEG) response to an external stimulus, reflect perceptual or cognitive demands. Other physiological measures include blood pressure variability (which is related to heart-rate variability) and electrodermal activity (see the reviews by Cain 2007; de Waard 1996; Wilson, Eggemeier 1991).

Changes in eye movement patterns can also be associated with attentional demands (Sheridan, Simpson 1979). May *et al.* (1990) reported that as task difficulty increased, the extent of spontaneous saccades (fast involuntary movements of the eye) decreased, whether the subjects were monitoring auditory or visual information. In a different experiment, done with pilots inside the cockpit, frequency of instrument fixation indicated its importance, while length of fixation was related to difficulty in interpreting the information (Wilson, Eggemeier 1991). Additional eye function measures are blink rate, duration and latency. One of the biggest advantages of using physiological measures is that they can be collected in real time. In some cases, however, they require specialized equipment and technicians, which may not

be available (de Waard 1996). Also, several researchers have criticized the fact that they are sensitive not only to stress caused by high workload, but to stress in general (Cain 2007).

*Performance-based measures*, as the name implies, consist in assessing mental workload through task performance. It is assumed that as workload increases, so do response times and errors. In addition, accuracy and number of completed tasks decrease (Huey, Wickens 1993). Therefore, it is possible to assess workload by tracking performance in a task with different difficulty levels. As mentioned earlier, however, findings like strategy adjustment, fatigue after-effects and speed-accuracy trade-offs have shown that it is possible for performance not to be affected, even when workload levels are high. Plus, performance errors can occur as often under high as under low workload. For this reason, a *secondary-task methodology* is often used. Here, a second task is performed at the same time and the effect of time-sharing on one of the tasks is measured. The goal of the second task is to use up the resources left over by the primary task, so that changes in performance are detected. If the emphasis is placed on the primary task and degradation measured on the secondary task, the technique is called *subsidiary-task paradigm*. In the *loading-task technique*, on the other hand, the emphasis is on the secondary task, with the degradation in performance measured in the primary task. According to Eggemeier and Wilson (1991), the most commonly used secondary tasks involve measurements of reaction-time, time estimation, memory search, and mental arithmetic tasks, among others.

Several of these paradigms have been successfully applied in aviation research to assess workload (e.g., Kantowitz *et al.* 1983; Schiflett 1980; Wickens *et al.* 1986). However, some researchers have criticized the use of such standard laboratory tasks, pointing out that they are strange to the operation environment under study (Lysaght *et al.* 1989). One alternative is to use embedded secondary tasks, that is, tasks that are usually performed during normal operations, but experimentally separable from the primary tasks. Examples of such tasks are aircraft radio communication activities, which were successfully used by Shingledecker (1980).

The final and most common type of methods are the *subjective methods*, which are based on the assumption that operators can reliably rate several aspects of the tasks. The advantages of these methods are that they are direct, easy to use, and inexpensive. Two of the most popular ones are the *NASA Task Load Index* or NASA-TLX and the *Subjective Workload Assessment Technique* or SWAT. The former requires people to rate the task from low to high on each of six scales: mental demand, physical demand, temporal demand, performance, effort

and frustration level. A weighting process takes the individual differences between the scales into account to compute an overall workload score (Hart, Staveland 1988). One common change made to NASA-TLX is to skip the weighting step and look at the average or sum of all ratings (Hart 2006). The latter method is divided into two phases: scale development and event scoring. In the scale development phase the individual is asked to rank-order 27 cards corresponding to three levels of three dimensions (time, mental effort and stress), which are then used to create an individual workload scale. The event scoring consists of the rating of the actual task situation using this scale (Reid *et al.* 1981).

Some disadvantages of the NASA-TLX and SWAT are their reliance on memory (usually these tests are performed after the tasks have been completed), their susceptibility to operators' bias and the task variability effect (people tend to use the whole rating scale, independently of the stimulus range). Finally, the use of emotional components in the NASA-TLX and SWAT tests have been criticized, due to the difficulty in relating the feelings of "frustration" and "anxiety" to the notion of allocated resources (Johnson, Proctor 2004). For this reason, some defend the use of unidimensional methods. One example of such a method is the Modified Cooper-Harper scale, in which the subjects go through a decision tree to arrive at a rating of mental workload (Cummings *et al.* 2006).

Other subjective methods are: the Subjective Workload Dominance Technique, the Visual, Auditory, Cognitive, Psychomotor method, the Workload Index method, the Multiple Resource Questionnaire, the Defence Research Agency Workload Scale, the Bedford scale, the Instantaneous Self-Assessment technique, among others (see the reviews by Cain 2007; de Waard 1996; Stanton *et al.* 2005).

Most of the subjective measures support only a relative evaluation, that is the experimental results require the definition and execution of a baseline scenario against which the results can be compared. In cases with considerable changes in the operational conditions, e.g., the elimination of radiotelephony communication in favour of a data link connection, a direct comparison of workload levels between the two solutions may not be possible, unless the definition of the baseline scenario covers such extraordinary conditions.

Choosing the right technique depends on the setting and the question of interest. Several factors should be taken in account when deciding on the best measure to use (Johnson, Proctor 2004; Eggemeier, Wilson 1991). First, the measure should be reliable in the sense that results must be replicated under similar conditions. Second, ease of use, which explains why subjective measures are so commonly used. Third, the measure must be sensitive to the changes in workload being manipulated.

Fourth, the chosen technique should be diagnostic, meaning that it needs to be able to identify the load imposed and on which resources. Fifth, intrusiveness level should also be considered, as the workload measure might cause a disruption in performance of the task under study. Finally, the measure needs to be accepted by the operators. If they find the whole experience boring, unpleasant or irrelevant, the data collected may not be valuable.

Different measures are sensitive to different aspects of workload. Part of the problem is the absence of a general definition of workload and of standardized procedures. In addition, the term workload is used to describe not only the demands imposed on the subject, but also the effort exerted to satisfy those demands and the consequences (physiological, subjective or performance) of those actions (Huey, Wickens 1993). Hence when selecting one measure, researchers need to consider whether the task under evaluation is predominantly psychomotor, cognitive or perceptual. Whenever possible, more than one method should be used, from different categories, and the advantages and disadvantages of each method carefully considered.

The question of how much workload is too much, is called the determination of a *workload redline* (Rueb, Vidulich 1994). Reid and Colle (1988), for example, attempted to find a critical SWAT score above which operators experienced performance problems and arrived at a value of  $40 \pm 10$ . However, according to de Waard (1996), by the time performance decreases, peak workload has already been crossed. Instead, he argues, workload redline should be the point after which the operator needs to exert some effort in order for performance to remain stable. This is also the point where stress begins if the situation lasts for some time. Lastly, individual differences on the amount of resources each person is willing, or capable, of allocating makes it very difficult to calculate a critical workload level.

Needless to say, the flight crew's perceived workload levels also depend on how well the user interface is designed. In general, pilots are satisfied with a system that helps them complete their tasks, while at the same time being easy to use without causing too much workload (Ahlstrom, Longo 2003). Any new interface which increases cognitive demand compared to current systems is poorly designed and may not be used to its full potential (disuse or misuse). Woods (1996) reported that experts cope with poorly designed automation by using only a few of the available functions, especially during high workload periods. This is ironic considering that most new functions in any automated system are usually introduced to support humans and reduce workload levels. Several other human-machine interface (HMI) aspects also need to be considered. For example, as pilots

might respond differently to a particular interface depending on workload levels, these will need to be tested under different conditions (that is, under abnormal, as well as under normal operations, in high and low workload).

The importance of managing workload in flight crews is reflected in the number of current regulations which refer specifically to it. For example, EASA CS 25.1322 specifies the effect of alarms and warnings in the pilots' workload (e.g. issue of false alarms, the use of too many colours in alerts) (EASA 2003). In the particular case of alert systems, a trade-off must be reached between falsely triggering an alarm and that of actual emergencies not being detected by the system (missed alarms). The consequences of each can be measured against the operator's workload to decide on where to place the criterion. For example, in very low workload levels, performance usually decreases and threats are sometimes overlooked. Thus, missed alarms can be very hazardous whereas false alarms may actually raise arousal levels and increase performance (Wilkinson 1964). Therefore, in these circumstances Endsley and Jones (2011) advise moving the alarm detection threshold in favour of more false alarms. The same suggestion is made during peak workload levels, because in order to cope with all tasks, the operator is forced to rely on automation and use his limited resources to monitor the system. Here, only the alerts raised by the system will be attended to, including false alarms, whereas missed alarms will probably not be detected by the operator. Finally, during normal workload levels the situation changes completely. In this case, false alarms are a source of annoyance to the operator who must attend to them, increasing workload to undesirable levels. Moreover, missed alarms may not be as hazardous since the operator is already attending to most systems and managing all tasks (Bustamante *et al.* 2004).

Alarms are designed to call attention to important events, especially in occasions when the operator is distracted or attending to other tasks. In theory, when an alarm sounds the information conveyed should be quickly understood and attended to with the proper action, without increasing workload levels. The easier it is for the pilots to detect and identify the auditory and visual signals, alerts, warnings, the lower the task demand and, hence, the lower the workload (Ahlstrom, Longo 2003). In reality, however, things are not so straightforward. In emergency situations it is now very common for several alerts to be activated at the same time forcing the operator to sort through them to understand the problem. In the Qantas Flight 32 incident in 2010 it took three crew members 50 minutes to process 54 failure messages. In addition, humans tend to seek for a confirmation of the problem before taking action, losing critical time. For

that reason, alarm systems should provide information that helps the operator identify the issue and quickly confirm the alarm's validity, without interfering with each other nor with the communication between the flight crew at critical times.

As reported above, high workload levels can lead to stress. The next section describes the causes of stress, and its physical and behavioural impacts.

### 3. Stress

Stress can be defined as the response of the body to stimuli that affect the normal physiological balance of a person, causing physical, mental or emotional strain. Hence, it is not a pathology per se, but an adaptation of the organism. A more recent approach is illustrated by Frankenhaeuser's bio-psychosocial model, which defines stress as the psychological and physiological reactions occurring when the individual perceives task demands as higher than the resources available to perform them (Martinussen, Hunter 2010). Therefore, what is considered stressful for one person, might not be for another. Examples of environmental demands are high workload, time constraints, conflicts and problems, whereas resources are determined by experience, personal abilities, physical and mental health, genetic factors and social support.

In some cases stress can be positive (eustress), if it focuses attention and increases arousal and vigilance levels. In this paper, however, I will focus on distress, caused by under- or over-arousal. Stimuli that cause stress are called stressors and they include physical factors such as noise, vibration, heat/cold, lighting, as well as social/psychological factors like anxiety, time pressure, mental load, fatigue, frustration, and anger Wickens *et al.* (2004). Common stressors in aircraft are: persistent radio communication noise, sudden alarms or warning horns, uncomfortable temperature (above 30°C or below 15°C), engine and system noise, vibration, cramped workspace, air quality, lighting conditions, etc. High workload levels can also cause stress when there is too much to do in too little time. Several aviation accidents have occurred because the crew was overloaded, which caused them to neglect key tasks Wickens *et al.* (2004). As just one example, in the accident with AA1420 in 1999, which crashed while attempting to land in bad weather, the pilots did not perform the last items of the "Before Landing" checklist.

Vigilance tasks combined with the need to maintain a high level of alertness can also cause stress. Thackray (1981) hypothesized that the monotony of a task inevitably leads to a decrease in arousal levels, requiring the operator to exert considerable effort to keep alert levels high enough to accomplish the task. A pilot

whose only task is to monitor automated systems in the cockpit often spends a considerable effort trying to remain alert and awake. Stress can also be triggered by a sudden onset of alerts and warnings in the cockpit. Harris (2004) described an incident report in which the pilot stated feeling so confused and frightened by the noise and flashing that his main priority was to cancel these out, instead of addressing what caused them to occur in the first place

Stress can be *acute* or *chronic*. As the names imply, acute stress is caused by an unexpected and sudden threat (e.g., engine failure), whereas chronic stress occurs when a stressor is present for a long time (e.g., relationship problems). Acute stress is the most common form of stress and it usually does not cause extensive physical or mental damage because of its short term duration. Chronic stress, on the other hand, is constant and persistent and can contribute to the development of several diseases. There are three stages of stress response. First, the body reacts by releasing hormones (*fight-or-flight response*), which gives rise to extra energy, muscular strength, and heightened hearing and vision (Alarm phase). That is followed by the Resistance phase, which corresponds to the body attempting to repair any damage caused by the stress, at the same time that it maintains the state of readiness. This long mobilization of body resources decreases resistance to other noxious stimuli. The final phase is Exhaustion, when the defence mechanisms collapse, leading to several symptoms and severe illness (Campbell, Bagshaw 2002). Physiological and psychological symptoms of stress include tachycardia, perspiration, muscular tension, insomnia, loss of appetite, headache, irritability, psychological disorders, gastro-intestinal diseases, muscular diseases, sexual disorders, cardio-vascular diseases, etc. (EUROCONTROL 1996).

Regarding flight crews, some recommendations have been made to help them cope with stress. In short, pilots are encouraged to avoid stress whenever possible by anticipating events and adopting a prevention strategy. On the one hand, if stress cannot be avoided (e.g. in the case of death of a family member), the symptoms should be recognized and the emphasis should be on the management of emotional reactions. If, on the other hand, there is a solution to the stress-causing problem, pilots should focus on solving it (Campbell, Bagshaw 2002). Some of the stress management strategies that pilots are encouraged to apply (when possible) include task delegating, handing over of aircraft control and taking a short break.

In general, stress affects how we perceive and process information, as well as what decisions we make, leading to an increase in the number of errors and mistakes and, thus, accidents/incidents. The most common

behaviour effects are (Martinussen, Hunter 2010; Stokes, Kite 1994):

- attentional narrowing or decrease in attention levels which translates into perceptual (narrower field of vision, selective hearing) and cognitive tunneling (attention is restricted to some cues, while others are not attended to);
- scattered and poorly organized visual scan;
- reductive thinking and filtering (considering only a few hypotheses, thus rejecting certain tasks or ignoring some warning signs);
- premature closure (making a decision without exploring all information);
- hurried decisions, even when there is no time pressure (leading to the speed-accuracy trade-off). Not surprisingly, the best decision-makers seem to be those who take their time under stress;
- regression: the crew might return to old procedures that may no longer be applicable, use non-standard phraseology when communicating, revert to the use of one's native language if different from the one being used (usually English) or look for items in a place where they used to be, but are no longer located;
- decrements in working memory capacity and retrieval;
- decrease in the ability to detect automation failures.

All of these may lead to an increase in the number of errors and, therefore, accidents/incidents (Martinussen, Hunter 2010; Stokes, Kite 1994).

The critical role of working memory in aviation cannot be overstated. At any given time while flying, a pilot needs to orient in a three-dimensional space (which requires temporary storage and manipulation of spatial and visual information) as well as keep in mind clearances, call-signs, instrument readings, advisories, briefings, and others - some of which are constantly being updated (Stokes, Kite 1994). Knowledge of regulations, systems and procedures (stored in long-term memory) is relatively resistant to stress, whereas more fundamental flying skills (the "stick and rudder control") can degrade significantly. A study by Wickens *et al.* (1988) using a computer-based simulation, investigated the effect of stress on pilot decision-making. The authors found that stress affected performance on scenarios that required the use of spatial working memory, but not on those that depended on the retrieval of knowledge stored in long-term memory. For this reason, experienced pilots, who rely more on long-term memory and on a rule-based approach to a problem, usually make less mistakes under stress than less experienced pilots. Another important issue is the relationship between stress and control. There is some evidence that stress reactions and changes in performance are less severe when individuals have

some control over the situation (Kantowitz, Casper 1988; Miller 1979). That is, when the warning signs are clear and a standard, trained solution to a problem can be applied, the situation is less stressful than if the signs are confusing and several messages need to be considered and processed at the same time. This finding has implications in the design of cockpit warning systems.

As mentioned previously, several physiological workload measures are also sensitive to general stress levels and can thus be used to measure the latter. Stressful situations lead to the release of hormones that can be quantified in blood, saliva and urine samples. Of interest are the catecholamines adrenaline and noradrenaline, as well as the steroid cortisol. Adrenaline levels have been associated with mental effort, whereas noradrenaline levels are determined by physical effort (Wilson, Eggemeier 1991). Catecholamine levels were found to depend on flight duration, level of experience, degree of responsibility and aircraft characteristics (Wilson, Eggemeier 1991). Kakimoto *et al.* (1988) reported that salivary cortisol levels increased more for pilots while they were in control of the aircraft than when they were not, especially during take-off and landing. Holmes and Rahe (1967) developed a scale that attempts to quantify stress. It consists of 43 stressful events that require personal adjustments (Scully *et al.* 2000), ranging from family matters, occupation, peer relationships, health, etc. Examples of events include a spouse's death (assigned 100 points), illness or injury (53 points) and trouble with the boss (23 points). A total score of 300 or more over a certain time period can have serious health consequences. Between 150 and 300 points leads to stress for about 50% of people. Under 150 points fewer than 30% of people become ill (EUROCONTROL 1996).

In terms of HMI, care should be taken to ensure that the interface does not induce stress, nor intensifies the psychological factors associated with it. For example, if the pilot is tired, a bad interface will not only be a source of stress, but will add to the pilot's fatigue, which leads to even more stress. Under stress humans tend to perform actions at the level of automatism and apply affordances. For example, a switch is usually off when it is moved backward. An overhead switch with the opposite result (move back to turn on) may still be properly used under low stress, but under high stress it has a high chance of leading to human error (Cardosi, Huntley 1993). This simple example highlights how HMI and stress can interact with negative consequences.

#### 4. Situation Awareness

Another very important concept associated with workload is Situation Awareness (SA). Like workload there is still some discussion in the literature about the definition of SA (is it the process of gaining awareness,

the product of awareness, or both?), what it refers to (knowledge stored in working memory or the result of information processing?), and whether it is a cognitive construct by itself or a term which congregates several different concepts (Salmon *et al.* 2009). Several researchers favour Endsley's definition that SA is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" or, put simply, "knowing what is going on" (Endsley 1995a). The first part of the definition, elements perception, consists mainly of information processing, the phase when 76% of all SA errors in pilots occur, according to Jones and Endsley (1996). At this level, the flight crew needs to attend to the airplane (all relevant lights, screens, indicators, etc.) and environmental cues. Cue salience and attentional narrowing, for example, can have a big impact on performance while engaged in this task of attention. The second SA level is the understanding of the situation in light of the operator's goals and objectives. The decision to abort take-off upon seeing a warning system alert is an example of this level of SA. Finally, in possession of that information the operator must predict the outcome of future events in order to make a decision regarding the best course of action. An example would be a pilot anticipating the trajectory of other airplanes. Fig. 3 shows a representation of each level of SA, as well as all factors that influence it.

Situation awareness requires paying attention to the relevant information, not only in the early information processing stages, but also during decision making and response execution. Consequently, in complex situations the limited capacity of attention can be quickly reached (Endsley 1995a). It is not possible for human operators to attend to all the relevant stimuli that are also processed at different speeds. Therefore, the amount of information that can be perceived at any time constitutes a bottleneck for SA (Endsley, Jones 2011).

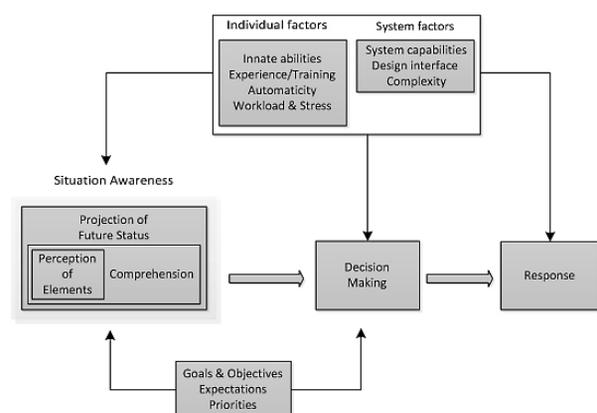


Fig. 3. Model of Situation Awareness (adapted from Endsley (1995a)).

Memory also plays an important role in SA, given that the individual needs to keep track of present and past events in the environment. According to Wickens (1984), the three levels of SA occur in *working memory*, suggesting a direct link between the two. In working memory the perceived stimuli are stored, processed and combined with previous knowledge in order to project what might happen in the future. Given its very limited capacity, it is the second bottleneck of SA (Endsley, Jones 2011).

*Long-term memory* is also fundamental to achieve SA. The access to stored information guides the expert operator on what to expect, freeing up working memory resources. Two of the mechanisms which help to circumvent the limitations of working memory are mental models and schemata. *Mental models* correspond to a systematic understanding of how something works and are the keys to levels 2 and 3 of SA (comprehension and projection). They also allow people to fill in the blanks when needed information is missing. *Schemata*, on the other hand, are stored representations of knowledge acquired from experience or vicariously through reading or hearing from others (Jones, Endsley 2000). They are like shortcuts that provide comprehension and projection in one step. The major difference between mental models and schemata is that only the latter are assumed to be stored and activated, whereas the former are thought to be creations of the moment arising from schemata (Wilson, Rutherford 1989).

The model of SA developed by Endsley also identifies two stages that proceed directly, and are separated, from SA: *decision making* and *response performance* (Fig. 3) (Endsley 1995a). A pilot can understand what is happening (i.e., have perfect SA), and yet not know what is the correct decision to make or how to execute it. Similarly, experienced pilots can make wrong decisions if they have an inaccurate SA. Finally, appropriate responses can in some cases be executed despite low SA, for example when a pilot cannot remember performing the act of extending the landing gear (Endsley 1988a, 1995a). Good decision making is usually, but not always, associated with expertise. An expert usually notices cues and patterns that might otherwise be missed by a novice (or is simply faster), and is able to anticipate a sequence of events (Sieck, Klein 2007).

Another construct involved in the acquisition of SA is the operators' specific *goals* for the tasks they are performing. These goals define how attention is directed and which elements in the environment are attended to. It also guides which mental models are chosen. This goal-driven process is also called *top-down information processing*. Pilots, for example, must attend to the elements in the cockpit which allow them to fly the aircraft to the destination. On the other hand, *data-driven*

or bottom-up processing occurs when information “catches” the operator’s attention independently of the goals (Endsley, Jones 2011). Examples are flashing lights or loud alarm noises, which might force the pilot to make immediate decisions based on the information provided by the instruments. As such, pilots need to be able to switch between data- and goal-driven processing while flying.

Also related to the goals are the operator’s *expectations* and *priorities*. These are based on mental models, experience, training, and communications from other sources (Endsley, Jones 2011). Preflight briefings, for example, are very important in defining a pilot’s expectations during flight. Like goals, expectations and priorities guide attention, providing a shortcut in processing the information without overloading working memory. One disadvantage is that they can lead to missing important cues when these are not expected to occur, for example a pilot could miss a ground proximity warning triggered by rising terrain during cruise flight. Goals, objectives, expectations and priorities affect not only SA, but decision making.

In a familiar situation, human behaviour is controlled by a set of already known and previously successful rules, whereas in a new situation, behaviour tends to be goal-controlled, as the subject accesses different mental representations, or schemas, from which the best one for the situation is chosen (Rasmussen 1983). One of the most well-known models of decision making is Rasmussen’s, which distinguishes between skill-, rule- or knowledge-based levels of behaviour (Rasmussen 1983):

*Skill-based* tasks are executed at an automated level. The entire motor process is executed without much conscious effort, attention or control. The actions are well trained and do not require a great deal of mental effort from the operator. Consequently, an operator is left with enough capacity to perform other tasks. An example is operating the brake if one wants to stop;

*Rule-based* tasks are well trained tasks, but they still require a bit of thinking. Therefore, they can be disturbed if the operator is distracted during execution. Proficient crew follow a script in which actions and reactions are expected to occur in a specific sequence. The stored rule goes: if this event occurs, then perform these actions. The boundary between skill-based and rule-based performance is not sharp, as it depends on the level of training and attention of the operator;

At the *knowledge-based* level the operators are confronted with relatively new situations, with things to choose and decide. As such they need to give mental attention to that task. The more knowledge-based tasks are, the more workload rises. Compared to machines, humans are relatively good performers at the knowledge-based level. This, however, creates a dilemma: the

level at which humans are better is the one that causes the most workload, but no machine can do it better.

For experienced pilots, hands-on flying (aviate) is mostly an automated motor skill (Gopher, Donchin 1986). Some navigation tasks, on the other hand, are rule-based, meaning that even though they are not automatic, they do not require high levels of mental resources, or that the pilot stops and thinks before acting. This is because the pilot has been trained on what to do, or has a lot of experience with such situations (Morris, Leung 2006). Finally, in new situations for which no rules exist it is necessary to resort to the slower, effortful and highly prone to error process of thinking things through (knowledge-based conceptual level). This level of demand also occurs when the pilots must handle multiple tasks and track the status of each one.

Other individual factors that affect the acquisition of SA are automaticity, experience and training. *Experience* and *training* lead to automaticity in mental processing, such that actions and behaviours become fast and effortless, freeing up mental resources. Through them people develop mental models, schemata and goal-directed processing (described above), which allow a high degree of time-sharing between tasks (Eggenmeier *et al.* 1991; Strayer, Kramer 1990). *Automaticity*, however, might lead to errors because events outside the routine are not attended to, e.g., if the pilots are given a different clearance than usual, they might not notice it and carry out the habitual action (Endsley 2010). This is one of the reasons why operations in the cockpit are highly scripted, with written procedures that detail the sequence of actions the pilots must take in each flight phase. Aircraft systems are usually set by memory and a checklist is used not only to make sure the most critical procedures are executed, but to protect against errors that occur when processes become too automatic. This approach has the advantage of allowing a large number of actions to be quickly and smoothly executed (Loukopoulos *et al.* 2003).

The amount of workload and stress also affects SA. Endsley (1995a) showed that SA and workload are independent constructs, with four possible combinations:

- low SA and low workload if the operator does not know what is happening and is not actively trying to find out;
- low SA and high workload if the operator is attending to too much information or too many tasks, thus not being able to process and integrate everything;
- high SA and low workload, in which the important information is being presented and correctly perceived and integrated (the ideal situation);
- high SA and high workload, when the operator is working hard, but successfully handling the situation.

System design, such as its capabilities, interface and complexity, also affect SA (Endsley 1995a). The way the information is presented to the pilot has a large impact on SA and workload. As avionics systems, flight management systems and other technology become more complex, the amount of data available increase. Consequently, there is also a raise in the number of components which need to be monitored, all of it potentially leading to high mental workload (Endsley 2010). The effects of system complexity can be alleviated by well-developed internal representations, which aids in directing attention and developing SA.

Endsley (2010) identified several general classes of elements that a pilot requires for SA. These are:

- geographical SA – location of own and other aircraft, terrain features, waypoints, and navigation fixes, climb/descent points; position relative to these, path to desired locations;
- spatial/temporal SA – attitude, altitude, heading, velocity, vertical velocity, G's, flight path; deviations from these and clearances;
- system SA – system status, functioning and settings; settings of radio, altimeter, and transponder equipment; ATC communications; flight modes and automation entries and settings; fuel; impact of malfunctions on system performance and flight safety;
- environmental SA – weather formations (area and altitudes affected); temperature, icing, ceilings, clouds, fog, sun, visibility, turbulence, winds, microbursts; instrument flight rules vs. visual flight rules conditions; areas and altitudes to avoid.

A more complete and detailed list of SA information requirements for commercial airline pilots can be found in Endsley *et al.* (1998).

Endsley and Jones (2011) and Endsley (1995b) suggested three different approaches to measure SA: process, performance, and direct measures. *Process measures* include verbal protocols and psychophysiological metrics. *Verbal protocols* consist of asking the subjects to “think out loud” while performing the task. The biggest disadvantages of this measure is its subjective nature and the reliance on the verbal skills of the subjects under study. The most common *psychophysiological methods* are eye movement recordings, EEG, and electrocardiogram. These are very useful in providing information on how attention is located, but handling the equipment and performing the data analyses can be difficult.

The second approach to study SA consists in using *performance measures*, which include *global measures of performance*, *external task measures* and *imbedded task measures* (Endsley 1995b). The first, as the name suggests, assesses SA based on overall performance in

realistic scenarios, such as the number of successful missions. The problem with this measure is the low diagnosticity and sensitivity, since performance can sometimes be masked by workload levels, decision-making, errors, etc. With external tasks measures, SA can be inferred by the way the subjects react to the manipulation or removal of information from the display. This is, however, a technique which is extremely intrusive and the experimenter might find it difficult to understand what the subjects know (or are aware of) based only on their overt behaviour. Another approach is to examine performance on specific subtasks (imbedded task measure), like deviations from a certain altitude. One disadvantage of using this method is that it might lead to misleading results if subjects focus exclusively on the subtask being evaluated to the exclusion of all others (Endsley 1995b).

*Direct measures* can be subjective or objective. *Subjective measures* include asking the subject or an observer to rate SA on some scale(s). One disadvantage of self-rating measures is that it assumes people can correctly evaluate their decisions and thoughts. The use of observers is also problematic since they need to infer SA based on what the subjects' knowledge of the situation is assumed to be. Examples of subjective tools are the SA Rating Technique (SART) and the SA Rating Scale (SARS) (Endsley, Jones 2011).

Like NASA-TLX, SART is a self-reporting, post-trial, multidimensional rating procedure designed for application in the aviation environment (Selcon *et al.* 1991). It includes ten independent dimensions further grouped into three domains: attentional demand, attentional supply, understanding. Both the 10-dimensional and the 3-dimensional versions can be implemented. See Selcon *et al.* (1991) for more information about this technique. SARS, on the other hand, consists of eight dimensions, covering 31 behaviours which are rated by the pilot on a six-point scale (Endsley, Jones 2011).

Finally, *objective measures* not only query the operators about aspects of the environment, but compare their responses with reality. Memory-probes are the most common objective measurement technique, reflecting the understanding that SA is the ability to keep track of what is going on (Vidulich 2002). One of the most popular objective measure of SA is the *SA Global Assessment Technique* (SAGAT), formulated by Endsley (1988b). It consists of running a scenario in an aircraft simulator, which at random times is halted so that a series of questions about the situation at that exact moment in time can be asked. At the end of the trials, the answers are evaluated based on what was happening in the simulation. According to Endsley (1988b), this comparison between the real and the perceived situation provides an objective measure of pilot SA. The greatest disadvantage of using this technique is the fact that the simulation needs to be stopped to collect the data.

The Situation Present Assessment Method, uses SAGAT-like queries while the operators perform the task (real-time). However, unlike SAGAT, it does not require a memory component: subjects with good SA should be able to answer the questions more quickly, because they can find the information faster (Durso *et al.* 1998). A variation of this technique is the SASHA, developed by EUROCONTROL to assess the SA of air traffic controllers (Stanton *et al.* 2005).

Even though in-depth research in SA started only in the 1980s, its importance in aviation cannot be underestimated. Endsley (1988a), for example, reported that the American Air Force considers a military pilot's ability to maintain SA as crucial to mission success and survivability. In recent years, several instruments were introduced in glass cockpit airplanes with the main goal of increasing SA. Two examples are the electronic horizontal situation indicator and the navigation display (Abbott *et al.* 1996). Another system that allows the pilots to maintain SA is the flight management system, which stores and displays information only when it is needed. In addition to that, pilots are able to keep workload at manageable levels by performing some tasks at a time of their choosing (Cardosi, Huntley 1993).

The importance of a good HMI design in establishing SA needs to be emphasized: only if the relevant information is presented in a clear and unambiguous manner are the pilots able to process and understand the situation, as well as anticipate future events. Following a sequence of accidents and incidents in the 90s attributed mostly to automation issues, the Federal Aviation Administration (FAA) commissioned a study of flight deck interfaces between the pilots and the automated systems of current generation aircraft (Abbott *et al.* 1996). The human factors team in charge of the report identified several issues associated with the flight crews' management of automation and situation awareness, such as:

- lack of understanding of automation's capabilities, limitations, modes, and operating principles and techniques;
- difficulty in choosing the appropriate automation level, or whether to turn automation on or off in non-normal situations;
- insufficient automation/mode awareness; and
- insufficient flight path awareness (especially terrain and aircraft energy state awareness).

The authors of the report made several recommendations to the FAA, among those the need to develop new ways to increase situation awareness, as well as to test interface designs for their susceptibility to hazardous states of awareness (e.g., underload, complacency, and fixation). Systems with several modes increase workload, introduce more possibilities for errors and can be difficult to learn. In most cases users must always be aware

of which mode is currently active, and understand how and when to use (and, more important, when not to use) each one. The autopilot system is an example of such a system. Several incidents and accidents have been reported that were caused by the pilots' misunderstanding of the autopilot mode (e.g., American Airlines flight AA903 incident; and TAROM flight ROT381).

Improved SA can usually be achieved by any perceptual display that helps the flight crew perceive, integrate and project the current situation into the future (Cardosi, Murphy 1995). In other words, displays should support SA at all times. Therefore, data should always be put in context to allow the pilot to immediately access the relevant information at the right time. But too much data can also be a problem. In order for the pilot to locate the needed information, clutter and everything that makes visual search difficult should be avoided (Ahlstrom, Longo 2003).

## 5. Automation

Even though the physical requirements of most jobs have decreased in the last 100 years, the cognitive or mental demands have increased dramatically (Johnson, Proctor 2004), and pilots are not an exception. Associated with that complexity is the need to multitask, which has become a common occurrence in our daily life. Ever since the Wright brothers' first powered flight in 1903 there has been an explosive growth in aircraft complexity. More and more instruments and gauges were introduced to increase safety or capability, but these also meant that pilots needed to spend an additional effort monitoring and controlling the aircraft. In the late 70s there were over 100 individual components in the cockpit, each providing one single piece of information, all of which needed to be monitored by the pilots. The visual complexity in the cockpit had reached a point where there was no more space for new instruments. Fortunately, new technology allowed for the introduction of digital displays, which show the same information in a smaller area (Curtis *et al.* 2010). Current systems consist of flexible multifunction displays, and the direct contact with control surfaces has been replaced by advanced computerized technologies. Among some of the decisions a pilot needs to make is when to attend to information (which is always available), where to look for it among all the different menu options, and how to interpret that information. Unfortunately, this trend toward more highly automated systems introduced new problems (Hollnagel 2012). Bainbridge called this outcome the *ironies of automation*, implying that automation may sometimes be more time consuming and/or incomprehensible than the manual operation of a system (Bainbridge 1983). The term *automation surprise* was introduced to designate all those occasions when humans

were left astonished and confused by the machine's behaviour.

Automation occurs whenever a task or action is executed by a machine, in most cases because it is more accurate and reliable than humans, and costs less (Parasuraman, Riley 1997). In complex systems, such as aircraft or power plants, automation may also help humans in the decision-making process, reducing workload and stress, and improving SA. This in turn translates into improved efficiency, flexibility and a decrease in the number of accidents. It is therefore expected that the trend towards extensive use of advanced automation will continue in the next decades. But even if automation made flying safer, more economic, reliable and comfortable, it also introduced a new class of issues associated with the loss of situation awareness and high workload.

In today's flight deck, maintaining an accurate SA is not an easy task and pilots report spending a lot of time working towards it, even after extensive use of the systems (Endsley, Jones 2011). In fact, the cause of several aircraft accidents has been attributed (in part) to lack of SA due to cockpit automation. This occurs because the pilots are *out-of-the-loop*, that is, they do not know what the system is doing or why. If an unexpected situation occurs that the system cannot handle and which requires the pilots to take over, they might misunderstand the information presented, miss important cues and thus be unable to properly evaluate the situation to decide on the proper action (Endsley, Jones 2011). As already mentioned in the paper, a specific case occurs when pilots assume that the system is working in one mode, when it is instead working in a different one. In cases like this, people tend to dismiss conflicting information (the *confirmation bias*) and may never realize the error they are making. This loss of mode awareness, which has already led to accidents, usually occurs when there is not enough feedback from the system to the human operator (Mouloua *et al.* 2010). As an example, the China Airlines Flight 140 crashed in Japan in 1994 while attempting to land, due to conflicting commands given by the pilots and the autopilot (set to Take-Off/Go-Around mode). In general, the less direct access the operator has to the system, the more important feedback is in order to maintain SA.

Two related factors that influence whether automation is used, and how, are trust and reliability. Too much trust leads to *complacency* (or *over-reliance*). In these cases automation is used when it should not be, or the human fails to actively monitor the machine due to the belief that nothing can go wrong, the *automation bias* (Parasuraman, Riley 1997). Errors resulting from this bias are generally split into *omission* and *commission* errors (Mosier *et al.* 1998). The former occurs when the operator does not execute the relevant action because it is not instructed by the machine. The latter is said to

take place when the wrong action is executed following erroneous information from the automated system. The opposite problem is the *disuse* of automation, when the operator is slow to respond or ignores automation, due to distrust or the assumption that the system is unreliable. This also occurs, for example, when there is an excessive number of false alarms, leading to the *cry wolf syndrome* in which warnings are ignored even when they should not be. One aircraft system which was victim of distrust when it was first introduced is the Ground Proximity Warning System, due to its high rate of false alarms (Wickens 2002). *Loss of proficiency or skill* has also been reported as a consequence of automation. For example, Wiener and Curry (1980) discuss the deficient flying skills of pilots who extensively use automatic equipment. Lee and Moray (1994) reported a direct relationship between self-confidence in manual skills and trust in automation: when trust exceeds self-confidence, automation is used. But the more often automation is used, the less trust the operator develops in her skills and the more impaired she becomes, which leads to further use of automation.

In general, humans are seen as being more flexible, adaptable and creative than machines, thus being better prepared to respond to unanticipated conditions. According to Miller and Parasuraman (2007), it is best to prevent either humans or automation to be exclusively in charge of the system. In other words, tasks should be shared between these two operators. This idea was first proposed by Sheridan and Verplank (1978) who characterized the levels of automation available in man-computer decision-making that can be manipulated with the goal of improving overall performance. These are presented in Table 1.

Table 1. Levels of Automation (adapted from Sheridan and Verplank (1978).

1	Computer provides no assistance
2	Computer provides options for actions
3	Computer offers options and suggests one, which humans need not follow
4	Computer suggests action, human accepts or rejects it and executes action
5	After selecting action, computer executes it if the human approves
6	Computer informs and allows humans time to stop action before automatic execution
7	Computer informs human after execution
8	Computer informs human after execution, if human asks
9	Computer informs human after execution if it, the computer, decides to.
10	Computer decides and acts autonomously – full automation

Automation is not an all-or-none concept, but should rather be seen as a continuum of levels (Parasuraman, Riley 1997). When deciding which level of automation to apply to a particular system, it is important to consider the implications of each and how they will affect performance and safety. Endsley and Kiris (1995) described a study where subjects had to take over manual operations after a failure in automation, and reported that subjects who previously performed the tasks with moderate and high levels of automation had lower SA and were slower than those subjects who had been performing the task manually all along. And most importantly, subjects working under full automation did not report lower workload levels before the failure, therefore supporting the view that system monitoring also causes workload. In a different experiment, Endsley and Kaber (1999) manipulated the level at which a task could be automated: display scanning, generation of options, option selection or implementation of actions. They found that during normal operations better performance occurred when humans generated and selected the options to be implemented by the machine, as opposed to when computers generated the options, or humans performed all four steps without assistance. But, as in the previous study, this last condition showed the best performance when automation failed. Parasuraman and Riley (1997) suggested that one way to solve this problem is to allow brief periods of manual task performance in between long periods of automation, which would also prevent the loss of skills linked to the limited opportunities to perform the task manually. These findings reinforce the notion that people seem to perform better when automation provides support at the level of collecting and presenting data, especially in routine, repetitive tasks, than at the higher cognitive levels (generating and selecting the best available options) (Endsley, Jones 2011).

One solution to some of the issues raised by automation is to develop a system where tasks are assigned to humans or automated systems in a flexible manner (*adaptive automation*). This implies that tasks are mostly performed by humans, and the switch to automation only occurs when they need support in order to meet the operational requirements (Rouse 1988). According to Moray *et al.* (2000), in time-critical situations final authority for decisions and actions should be allocated to the machine. Once tasks become more manageable, automation should be deactivated and the operator takes over. Supporters of this technology assert that in situations when humans need assistance they might not have the resources available to realize it and request support from the system (Rouse 1988). In fact, in several aircraft and nuclear power plants systems are already at level 7 or higher (see Table 1).

Several methods have been proposed by which adaptive automation is to be implemented, including the occurrence of specific events, performance measurements, physiological assessment (e.g., pupilometry, heart rate variability, EEG) (Bailey *et al.* 2006; Parasuraman *et al.* 1992; Wilson, Russell 2007). Prinzel *et al.* (2003), for example, used EEG measurements to switch between automatic and manual task modes – when the operator's engagement level increased above a baseline condition, the task would switch to automatic mode. Below this value, the manual mode would remain engaged. Subjects in the adaptive automation group performed better and reported lower workload levels than those in the two control groups. In a different study, Kaber and Riley (1999) used performance in a secondary task to direct or suggest to subjects to use automation in the primary task. The authors reported that performance was better in all tasks for the mandated group than for the non-mandated participants. With the development of methods to determine optimal allocation strategies, adaptive automation has been found to reduce workload and improve situation awareness (Bailey *et al.* 2003; Kaber, Riley 1999). Despite these promising results, such automation systems have not yet been considered a team player by human operators and need to be designed with great caution. Problems can arise if, for example, the system's behaviour is unpredictable to the operator (Billings, Woods 1994) or if it is implemented unexpectedly or at a time the user does not wish it (Ahlstrom, Longo 2003).

Research by Kaber *et al.* (2005) suggests that adaptive automation is more effective when applied at those levels where automation is found to provide better support, that is, at information acquisition and action implementation, rather than at analysis and decision making. The main disadvantages of adaptive automation have already been mentioned previously when discussing the disadvantages of automation in general: reduced SA; mistrust, complacency and over-reliance; loss of skills and performance, etc. (Miller *et al.* 2005).

Miller and Parasuraman (2007) defend another approach. The authors suggest that automation should be *adaptable*, that is, humans remain in charge and decide how much automation to use. The authors recognize that adaptable automation increases cognitive demand as the operators are responsible for supervising the system and decide on the level of automation, but defend that this is preferable to adaptive automation, which decreases workload by decreasing user involvement and responsibility. As mentioned before, however, humans might not be the best judges. A study by Bailey *et al.* (2006) indicated that subjects who had the possibility to switch to automated control still preferred manual control, despite reporting higher workload levels than participants in the adaptive condition. Moreover, this reluctance to use

automation had a detrimental impact on performance. On the other hand, Kidwell *et al.* (2012) compared the effect of the two types of automation and concluded that adaptable automation not only increased the probability of the operator detecting changes in the system, but also increased confidence in decision-making, by keeping the operator in-the-loop and alert to unexpected events. Thus, it is expected that the detrimental effects of automation are mitigated in adaptable automation (Miller *et al.* 2005).

Independent of the level or type of automation, one of the most crucial parts in the development of any system is the design of the *human-machine interface*. New avionics and cockpit displays need to help the pilots in building a correct and timely assessment of the situation, not only during normal flight conditions but also in case of malfunctions or emergencies. As already mentioned, under high workload and stress humans are highly susceptible to the confirmation bias and cognitive tunneling. Hence, it is exactly in those moments that good cockpit design can positively impact human performance (Cardosi, Huntley 1993). The HMI design should then support the acquisition of SA, help the pilots focus their attention on the appropriate displays and messages, prevent human error, avoid distractions from other sources, and compensate for the cognitive tunnel (Cardosi, Murphy 1995). This should be done without causing a cognitive overload, which requires a robust human-machine interaction philosophy.

One way of evaluating the HMI of a system is by looking at its usability, mostly by collecting data of users interacting with the interface. Usability relates to the perceptual and physical aspects of the interface (display formatting, graphics, feedback, etc.) and their effectiveness for the user to achieve his goals (Nielsen 1993). According to Nielsen (1993), the usability of a system can have several components or features, the most common being:

- learnability (system should be easy to learn and the solutions intuitive);
- efficiency (the system enables the task to be completed in a timely, effective and economical fashion);
- memorability (what is learned should be easy to remember);
- low error rate (users should make few errors and easily recover from errors that will eventually be made);
- satisfaction (users should like to use it).

To improve a systems' usability, the concept of user-centred design, has been proposed by a number of researchers. A user-centred design of the HMI focuses on the human operator in order to reduce the number of errors and increase efficiency, as opposed to

a technology-centred approach, which focuses exclusively on the technological feasibility without considering the impact on the human operator. Three principles for user-centred design were described by Endsley and Jones (2011). These are:

- organize technology around the user's goals, tasks, and abilities. Systems should not exceed the operators' mental, perceptual and physical capabilities. Additionally, the design of complex systems should support the changing goals and tasks of the user operating in a dynamic environment;
- technology should be organized around the way operators process information and make decisions. In complex real-world settings, humans spend a considerable amount of time assessing a situation and comparing it with information stored in long-term memory, before choosing an action. Therefore, the system should help the user maintain situation awareness;
- technology must keep the user in control and aware of the state of the system. The system cannot put the human operator out-of-the-loop, which would prevent system monitoring and the maintenance of situation awareness.

## 6. Conclusion

The expected increase in the number of aircraft and passengers over the next decades will lead to a higher volume of traffic in the skies, which will inevitably result in more fatal accidents if today's accident rate is not further reduced. In addition, major system changes brought by SESAR and NextGen, growing economic pressure and greater global environmental consciousness demand better and more reliable technology in the cockpit, which usually translate into increased automation. Unfortunately recent accidents and incidents attributed to human error have highlighted the need to examine the role of automation and pilots' position as managers and supervisors, especially in periods of high workload and stress.

Workload can be operationally defined in terms of the memory load imposed by the system on the pilot, the number of mental transformations of data that the system requires, or how fast and accurately the flying task is performed (Cardosi, Murphy 1995). Closely associated with workload is stress, defined as a state of physical, mental or emotional strain caused by internal or external factors that are perceived as threatening. As shown in this paper, these are fundamental concepts in aviation psychology, which cannot be dissociated from situation awareness: the pilots' understanding of what is going on with the airplane, its systems and the conditions outside the aircraft (e.g. weather degradation). Workload

reduction, for example, should never be achieved at the expense of a loss of situation awareness.

High workload, stress and low situation awareness, all increase the probability of the pilot making a mistake. Between 60 to 70% of all accidents involving commercial aircraft have been attributed to human error (Dismukes *et al.* 2007). As air traffic increases and systems become more complex, it is crucial to understand what causes those accidents and how to reduce their number. Most researchers agree, however, that usually there is no single cause for accidents. The majority are a result of a conjunction of failures, as illustrated by Reason's latent failure model. According to Reason (1990), in every system there are barriers or layers in place to prevent accidents. As long as the barriers are intact, there are no accidents. Unfortunately layers incorporate weaknesses or even have failures, represented as holes. When events occur in such a way that each barrier of protection fails, then the holes are aligned and an accident is waiting to happen (Martinussen, Hunter 2010).

Human error should not be considered an exceptional event, but rather a component of complex systems which cannot be completely eliminated and therefore should be seen as inevitable and normal. Their frequency can be reduced by improved design, better training, redundancy, but they will not go away (Perrow 1984). As seen above, simply increasing automation levels does not eliminate human error. Changes in technology introduce new challenges and tasks (which require additional knowledge, attentional demands, procedures), creating the potential for different types of errors and new ways a system can break down.

### Funding

The research leading to these results has received funding from the European Commission's Seventh Framework Programme FP7/2012-2016 [grant number ACP2-GA-2012-314501]. Portions of this paper were included in the delivery "Scope and State-of-the-Art" of ACROSS. The information and views set out in this paper are those of the author and do not necessarily reflect the official opinion of the ACROSS Consortium or the European Commission.

### Disclosure statement

The author declares no competing financial, professional, or personal interests from other parties.

### References

- Abbott, K.; Slotte, S. M.; Stimson, D. K. 1996. *Federal Aviation Administration Human Factors Team Report: The Interfaces Between Flightcrews and Modern Flight Deck Systems*. Washington, DC: Aircraft Certification Service, Federal Aviation Administration.
- Ahlstrom, V.; Longo, K. 2003. *Human Factors Design Standard for Acquisition of Commercial-Off-The-Shelf Subsystems, Non-Developmental Items, and Developmental Systems (DOT/FAA/CT-03/05)*. Atlantic City, NJ: Federal Aviation Administration Technical Center.
- Airbus. 2015. *Global Market Forecast 2015-2034* [online], [cited 9 May 2016]. 49 p. Available from Internet: <http://www.airbus.com/company/market/forecast/>
- Bailey, N. R.; Scerbo, M. W.; Freeman, F. G.; Mikulka, P. J.; Scott, L. A. 2003. A brain-based adaptive automation system and situation awareness: The role of complacency potential, *Proceedings of the Human Factors and Ergonomics Society* 47: 1048–1052. <http://dx.doi.org/10.1177/154193120304700901>
- Bailey, N. R.; Scerbo, M. W.; Freeman, F. G.; Mikulka, P. J.; Scott, L. A. 2006. A comparison of a brain-based adaptive system and a manual adaptable system for invoking automation, *Human Factors* 48(4): 693–709. <http://dx.doi.org/10.1518/001872006779166280>
- Bainbridge, L. 1983. Ironies of automation, *Automatica*, 19(6): 775–779. [http://dx.doi.org/10.1016/0005-1098\(83\)90046-8](http://dx.doi.org/10.1016/0005-1098(83)90046-8)
- Beatty, J. 1982. Phasic not tonic pupillary responses vary with auditory vigilance performance, *Psychophysiology* 19(2): 167–172. <http://dx.doi.org/10.1111/j.1469-8986.1982.tb02540.x>
- Billings, C. E.; Woods, D. D. 1994. Concerns about adaptive automation in aviation systems, in R. Parasuraman, M. Mouloua (Eds.). *Human Performance in Automated Systems: Current Research and Trends*. Hillsdale, NJ: Erlbaum, 264–269
- Boeing. 2014. *Statistical Summary of Commercial Jet Airplane Accidents. Worldwide Operations 1959-2014* [online], [cited 9 May 2016]. 26 p. Available from Internet: <http://www.boeing.com/news/techissues/pdf/statsum.pdf>
- Boeing. 2015. *Current Market Outlook 2015-2034* [online], [cited 9 May 2016]. 52 p. Available from Internet: <http://www.boeing.com/commercial/market/>
- Bustamante, E. A.; Anderson, B. L.; Bliss, J. P. 2004. Effects of varying the threshold of alarm systems and task complexity on human performance and perceived workload, *Proceedings of the Human Factors and Ergonomics Society* 48: 1948–1952. <http://dx.doi.org/10.1177/154193120404801633>
- Cain, B. 2007. *A review of the mental workload literature (NATO Report No. RTO-TR-HFM-121-Part-II)*. Toronto, Canada: Defence Research and Development.
- Campbell, R. D.; Bagshaw, M. 2002. *Human performance and limitations in aviation*. 3rd ed. Oxford: Blackwell Science. <http://dx.doi.org/10.1002/9780470774472>
- Cardosi, K. M.; Huntley, M. S. 1993. *Human Factors for Flight Deck Certification Personnel (DOT/FAA/RD-93/5)*. Cambridge, MA: U.S. Department of Transportation.
- Cardosi, K. M.; Murphy, E. D. 1995. *Human Factors in the Design and Evaluation of Air Traffic Control Systems (DOT/FAA/RD-95/3)*. Cambridge, MA: U.S. Department of Transportation.
- Casali, J. G.; Wierwille, W. W. 1983. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load, *Human Factors* 25(6): 623–641. <http://dx.doi.org/10.1177/001872088302500602>
- Curtis, M. T.; Jentsch, F.; Wise, J. A. 2010. Aviation Displays, in E. Salas, D. Maurino (Eds.). *Human Factors in Aviation*. Amsterdam, The Netherlands: Academic Press/Elsevier, 439–478. <http://dx.doi.org/10.1016/B978-0-12-374518-7.00014-6>

- Darby, R. 2006. Commercial Jet Hull Losses, Fatalities Rose Sharply in 2005, *Aviation Safety World*, 51–53.
- de Waard, D. 1996. *The measurement of drivers' mental workload*. PhD thesis, University of Groningen. [online], [cited 9 May 2016]. Available from Internet: <http://www.home.zonnet.nl/waard/mwl.htm>
- Diamond, D. M.; Campbell, A. M.; Park, C. R.; Halonen, J.; Zoladz, P. R. 2007. The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law, *Neural plasticity* 2007: 1–33. <http://dx.doi.org/10.1155/2007/60803>
- Dismukes, R. K.; Berman, B. A.; Loukopoulos, L. D. 2007. *The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents*. Aldershot, England: Ashgate.
- Dismukes, R. K.; Nowinski, J. 2006. Prospective Memory, Concurrent Task Management, and Pilot Error, in A. Kramer, D. Wiegmann, A. Kirlik (Eds.). *Attention: From Theory to Practice*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195305722.003.0016>
- Durso, F. T.; Hackworth, C. A.; Truitt, T. R.; Crutchfield, J.; Nikolic, D.; Manning, C. A. 1998. Situation awareness as a predictor of performance for en-route air traffic controllers, *Air Traffic Control Quarterly* 6(1): 1–20.
- Easterbrook, J. A. 1959. The effect of emotion on cue utilization and the organization of behavior, *Psychological Review* 66(3): 183–201. <http://dx.doi.org/10.1037/h0047707>
- Eggemeier, F. T.; Wilson, G. F. 1991. Performance-based and subjective assessment of workload in multi-task environments, in D. L. Damos (Ed.). *Multiple-task performance*. Boca Raton, FL: CRC Press, 217–278.
- Eggemeier, F. T.; Wilson, G. F.; Kramer, A. F.; Damos, D. L. 1991. Workload assessment in multi-task environments, in D. L. Damos (Ed.). *Multiple-task performance*. Boca Raton, FL: CRC Press, 207–216.
- Endsley, M. R. 1988a. Design and evaluation for situation awareness enhancement, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 32(2): 97–101. <http://dx.doi.org/10.1109/naecon.1988.195097>
- Endsley, M. R. 1988b. Situation awareness global assessment technique (SAGAT), in *Proceedings of the National Aerospace and Electronics Conference*. New York, NY: IEEE, 789–795. <http://dx.doi.org/10.1109/naecon.1988.195097>
- Endsley, M. R. 1995a. Toward a theory of situation awareness in dynamic systems, *Human Factors* 37(1): 32–64. <http://dx.doi.org/10.1518/001872095779049543>
- Endsley, M. R. 1995b. Measurement of situation awareness in dynamic systems, *Human Factors* 37(1): 65–84. <http://dx.doi.org/10.1518/001872095779049499>
- Endsley, M. R. 2010. Situation Awareness in Aviation Systems, in J. A. Wise, V. D. Hopkin, D. J. Garland. (Eds.). *Handbook of aviation human factors*. 2nd ed. Boca Raton, FL: CRC Press, 12.1–12.22.
- Endsley, M. R.; Farley, T. C.; Jones, W. M.; Midkiff, A. H.; Hansman, R. J. 1998. *Situation Awareness Information Requirements for Commercial Airline Pilots (ICAT-98-1)*. Cambridge, MA: Massachusetts Institute of Technology International Center for Air Transportation.
- Endsley, M. R.; Jones, D. G. 2011. *Designing for situation awareness: an approach to user-centered design*. 2nd ed. Boca Raton, FL: CRC Press. <http://dx.doi.org/10.1201/b11371>
- Endsley, M. R.; Kaber, D. B. 1999. Level of automation effects on performance, situation awareness and workload in a dynamic control task, *Ergonomics* 42: 462–492. <http://dx.doi.org/10.1080/001401399185595>
- Endsley, M. R.; Kiris, E. O. 1995. The out-of-the-loop performance problem and level of control in automation, *Human Factors* 37: 381–394. <http://dx.doi.org/10.1518/001872095779064555>
- EASA. 2003. *Certification Specifications and Acceptable Means of Compliance, for Large Aeroplanes (CS-25)*. Cologne, Germany: European Aviation Safety Agency.
- EUROCONTROL. 1996. *Human Factors Module – Stress (HUM.ET1.ST13.2000-REP-01)*.
- Gladstones, W. H.; Regan, M. A.; Lee, R. B. 1989. Division of attention: The single-channel hypothesis revisited, *The Quarterly Journal of Experimental Psychology* 41(1): 1–17. <http://dx.doi.org/10.1080/14640748908402350>
- Gopher, D.; Braune, R. 1984. On the psychophysics of workload: Why bother with subjective measures? *Human Factors* 26(5): 519–532.
- Gopher, D.; Donchin, E. 1986. Workload: An examination of the concept, in L. Kaufman, K. Boff (Eds.). *Handbook of perception and human performance*, vol. 2. New York, NY: Wiley, 41.1–41.49.
- Hancock, P.; Warm, J. 1989. A dynamic model of stress and sustained attention, *Human Factors* 31(5): 519–537. doi: 10.1177/001872088903100503
- Harris, D. (Ed.). 2004. *Human factors for civil flight deck design*. Aldershot, UK: Ashgate
- Hart, S. G.; Staveland, L. E. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, *Human Mental Workload* 1(3): 139–183. [http://dx.doi.org/10.1016/S0166-4115\(08\)62386-9](http://dx.doi.org/10.1016/S0166-4115(08)62386-9)
- Hart, S. G. 2006. NASA-task load index (NASA-TLX); 20 years later, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50(9): 904–908. <http://dx.doi.org/10.1177/154193120605000909>
- Hollnagel, E. 2012. Coping with complexity: past, present and future, *Cognition, Technology & Work* 14(3): 199–205. <http://dx.doi.org/10.1007/s10111-011-0202-7>
- Holmes, T. H.; Rahe, R. H. 1967. The social readjustment rating scale, *Journal of psychosomatic research* 11(2): 213–218. [http://dx.doi.org/10.1016/0022-3999\(67\)90010-4](http://dx.doi.org/10.1016/0022-3999(67)90010-4)
- Huey, F. M.; Wickens, C. D. 1993. *Workload transition: Implications for individual and team performance*. Washington, DC: National Academy Press.
- Johnson, A.; Proctor, R. W. 2004. *Attention: Theory and practice*. Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781483328768>
- Jones, D. G.; Endsley, M. R. 1996. Sources of situation awareness errors in aviation, *Aviation, Space, and Environmental Medicine* 67(6): 507–512.
- Jones, D. G.; Endsley, M. R. 2000. Overcoming representational errors in complex environments, *Human Factors* 42(3): 367–378. <http://dx.doi.org/10.1518/001872000779698187>
- Kaber, D. B.; Riley, J. 1999. Adaptive automation of a dynamic control task based on secondary task workload measurement, *International Journal of Cognitive Ergonomics* 3(3): 169–187. [http://dx.doi.org/10.1207/s15327566ijce0303\\_1](http://dx.doi.org/10.1207/s15327566ijce0303_1)
- Kaber, D. B.; Wright, M. C.; Prinzel, L. J.; Clamann, M. P. 2005. Adaptive automation of human-machine system information-processing functions, *Human Factors* 47(4): 730–741. <http://dx.doi.org/10.1518/001872005775570989>
- Kakimoto Y. U. K. I. K. O.; Nakamura A. K. I. O.; Tarui H. I. D. E. O.; Nagasawa, Y. U. K. O.; Yagura, S. H. I. G. E. Y. U. K. I. 1988. Crew workload in JASDF C-1 transport flights: I. Change in heart rate and salivary cortisol, *Aviation, Space, and Environmental Medicine* 59(6): 511–516.

- Kantowitz, B. H.; Casper, P. A. 1988. Human Workload in Aviation, in E. Wiener, D. Nagel (Eds.). *Human Factors in Aviation*. S. Diego, CA: Academic Press, 157–188. <http://dx.doi.org/10.1016/b978-0-08-057090-7.50012-6>
- Kantowitz, B. H.; Hart, S. G.; Bortolussi, M. R. 1983. Measuring pilot workload in a moving-base simulator: I. Asynchronous secondary choice-reaction task, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 27(4): 319–322. <http://dx.doi.org/10.1177/154193128302700413>
- Kantowitz, B. H.; Knight, J. L. 1976. Testing tapping timesharing, II: auditory secondary task, *Acta Psychologica* 40(5): 343–362. [http://dx.doi.org/10.1016/0001-6918\(76\)90016-0](http://dx.doi.org/10.1016/0001-6918(76)90016-0)
- Kidwell, B.; Calhoun, G. L.; Ruff, H. A.; Parasuraman, R. 2012. Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56(1): 428–432. <http://dx.doi.org/10.1177/1071181312561096>
- Latorella, K. A. 1999. *Investigating interruptions: Implications for flightdeck performance* (NASA/TM-1999-209707). Hampton, VA: National Aeronautics and Space Administration.
- Lee, J. D.; Moray, N. 1994. Trust, self-confidence, and operators' adaptation to automation, *International Journal of Human-Computer Studies* 40(1): 153–184. <http://dx.doi.org/10.1006/ijhc.1994.1007>
- Loukopoulos, L. D.; Dismukes, R. K.; Barshi, I. 2003. Concurrent task demands in the cockpit: Challenges and vulnerabilities in routine flight operations, in *Proceedings of the 12th International Symposium on Aviation Psychology*. Dayton, OH: Wright State University Press, 737–742
- Lysaght, R. J.; Hill, S. G.; Dick, A. O.; Plamondon, B. D.; Linton, P. M.; Wierwille, W. W.; Zaklad, A. L.; Bittner Jr.; A. C.; Wherry, R. J. 1989. *Operator workload: Comprehensive review and evaluation of operator workload methodologies* (No. TR-2075-3). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.
- Mackworth, N. H. 1948. The breakdown of vigilance during prolonged visual search, *The Quarterly Journal of Experimental Psychology* 1(1): 6–21. <http://dx.doi.org/10.1080/17470214808416738>
- Martinussen, M.; Hunter, D. R. 2010. *Aviation psychology and human factors*. Boca Raton, FL: CRC Press.
- May, J. G.; Kennedy, R. S.; Williams, M. C.; Dunlap, W. P.; Brannan, J. R. 1990. Eye movement indices of mental workload, *Acta Psychologica* 75(1): 75–89. [http://dx.doi.org/10.1016/0001-6918\(90\)90067-P](http://dx.doi.org/10.1016/0001-6918(90)90067-P)
- McLeod, P. 1977. A dual task response modality effect: Support for multiprocessor models of attention, *The Quarterly Journal of Experimental Psychology* 29(4): 651–667. <http://dx.doi.org/10.1080/14640747708400639>
- Meister, D.; Gawron, V. 2010. Measurement in Aviation Systems, in J. A. Wise, D. Hopkin, D. J. Garland (Eds.). *Handbook of aviation human factors*. 2nd ed. Boca Raton, FL: CRC Press, 3.1–3.15.
- Miller, C. A.; Funk, H.; Goldman, R.; Meisner, J.; Wu, P. 2005. Implications of adaptive vs. adaptable UIs on decision making: Why “automated adaptiveness” is not always the right answer, in *Proceedings of the 1st International Conference on Augmented Cognition*, 22–27 July 2005, Las Vegas, NV.
- Miller, C. A.; Parasuraman, R. 2007. Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control, *Human Factors* 49(1): 57–75. <http://dx.doi.org/10.1518/001872007779598037>
- Miller, S. M. 1979. Controllability and human stress: Method, evidence and theory, *Behaviour Research and Therapy* 17(4): 287–304. [http://dx.doi.org/10.1016/0005-7967\(79\)90001-9](http://dx.doi.org/10.1016/0005-7967(79)90001-9)
- Moray, N.; Inagaki, T.; Itoh, M. 2000. Adaptive automation, trust, and self-confidence in fault management of time-critical tasks, *Journal of Experimental Psychology: Applied* 6(1): 44–58. <http://dx.doi.org/10.1037/1076-898X.6.1.44>
- Morris C. H.; Leung Y. K. 2006. Pilot mental workload: how well do pilots really perform? *Ergonomics* 49(15): 1581–1596. <http://dx.doi.org/10.1080/00140130600857987>
- Mosier, K. L.; Skitka, L. J.; Heers, S.; Burdick, M. 1998. Automation bias: Decision making and performance in high-tech cockpits, *The International Journal of Aviation Psychology* 8(1): 47–63. [http://dx.doi.org/10.1207/s15327108i-jap0801\\_3](http://dx.doi.org/10.1207/s15327108i-jap0801_3)
- Mouloua, M.; Hancock, P.; Jones, L.; Vincenzi, D. 2010. Automation in Aviation Systems: Issues and Considerations, in J. A. Wise, D. Hopkins, D. J. Garland (Eds.). *Handbook of aviation human factors*. 2nd ed. Boca Raton, FL: CRC Press.
- Nielsen, J. 1993. *Usability Engineering*. San Diego, CA: Academic Press/Morgan Kaufman.
- Parasuraman, R.; Bahri, T.; Deaton, J. E.; Morrison, J. G.; Barnes, M. 1992. *Theory and design of adaptive automation in aviation systems. Progress Report (AD-A254 595)*. Washington, DC: Cognitive Science Laboratory, The Catholic University of America.
- Parasuraman, R.; Mouloua, M. 1987. Interaction of signal discriminability and task type in vigilance decrement, *Perception & Psychophysics* 41(1): 17–22. <http://dx.doi.org/10.3758/BF03208208>
- Parasuraman, R.; Riley, V. 1997. Humans and automation: Use, misuse, disuse, abuse, *Human Factors* 39(2): 230–253. <http://dx.doi.org/10.1518/00187209778543886>
- Perrow, C. 1984. *Normal Accidents: Living with High Risk Technologies (Updated)*. Princeton, NJ: Princeton University Press.
- Prinzel, L. J.; Freeman, F. G.; Scerbo, M. W.; Mikulka, P. J.; Pope, A. T. 2003. Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component, *Human Factors* 45(4): 601–613. <http://dx.doi.org/10.1518/hfes.45.4.601.27092>
- Rasmussen, J. 1983. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models, *IEEE Transactions on Systems, Man and Cybernetics* 13(3): 257–266. <http://dx.doi.org/10.1109/TSMC.1983.6313160>
- Reason, J. 1990. The contribution of latent human failures to the breakdown of complex systems, *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 327(1241): 475–484. <http://dx.doi.org/10.1098/rstb.1990.0090>
- Reid, G. B.; Colle, H. A. 1988. Critical SWAT values for predicting operator overload, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 32(19): 1414–1418.
- Reid, G. B.; Shingledecker, C. A.; Eggemeier, F. T. 1981. Application of conjoint measurement to workload scale development, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 25(1): 522–526. <http://dx.doi.org/10.1177/1071181381025001134>

- Rouse, W. B. 1988. Adaptive aiding for human/computer control, *Human Factors* 30(4): 431–438. <http://dx.doi.org/10.1177/001872088803000405>
- Rueb, J. D.; Vidulich, M. A.; Hassoun, J. A. 1994. Use of Workload Redlines: A KC-135 Crew-Reduction Application, *The International Journal of Aviation Psychology* 4(1): 47–64. [http://dx.doi.org/10.1207/s15327108ijap0401\\_3](http://dx.doi.org/10.1207/s15327108ijap0401_3)
- Salmon, P. M.; Stanton, N. A.; Walker, G. H.; Jenkins, D.; Ladva, D.; Rafferty, L.; Young, M. 2009. Measuring Situation Awareness in complex systems: Comparison of measures study, *International Journal of Industrial Ergonomics* 39(3): 490–500. <http://dx.doi.org/10.1016/j.ergon.2008.10.010>
- Schellekens, J. M.; Sijtsma, G. J.; Vegter, E.; Meijman, T. F. 2000. Immediate and delayed after-effects of long lasting mentally demanding work, *Biological Psychology* 53(1): 37–56. [http://dx.doi.org/10.1016/S0301-0511\(00\)00039-9](http://dx.doi.org/10.1016/S0301-0511(00)00039-9)
- Schifflett, S. G. 1980. *Evaluation of a Pilot Workload Assessment Device to Test Alternate Display Formats and Control Handling Qualities (Report Nr. ADA087604)*. Patuxent River, MD: Naval Air Test Center.
- Schumacher, E. H.; Seymour, T. L.; Glass, J. M.; Fencsik, D. E.; Lauber, E. J.; Kieras, D. E.; Meyer, D. E. 2001. Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck, *Psychological Science* 12(2): 101–108. <http://dx.doi.org/10.1111/1467-9280.00318>
- Scully, J. A.; Tosi, H.; Banning, K. 2000. Life event checklists: Revisiting the social readjustment rating scale after 30 years, *Educational and psychological measurement* 60(6): 864–876. <http://dx.doi.org/10.1177/00131640021970952>
- Selcon, S. J.; Taylor, R. M.; Koritsas, E. 1991. Workload or situational awareness?: TLX vs. SART for aerospace systems design evaluation, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 35: 62–66. <http://dx.doi.org/10.1518/107118191786755706>
- Sheridan, T. B.; Simpson R. W. 1979. *Toward the definition and measurement of the mental workload of transport pilots (Flight Transportation Laboratory Report R 79-4)*. Cambridge, MA.: Massachusetts Institute of Technology.
- Sheridan, T. B.; Verplank, W. L. 1978. *Human and computer control of undersea teleoperators*. Cambridge, MA: MIT Man-Machine Laboratory.
- Shingledecker, C. A. 1980. Enhancing Operator Acceptance and Noninterference in Secondary Task Measures of Workload, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 24(1): 674–677.
- Sieck, W. R.; Klein, G. 2007. Decision-Making, in T. F. Durso, R. S. Nickerson, S. Dumais, S. Lewandowsky, T. Perfect (Eds.), *Handbook of Applied Cognition*. 2nd ed. Chichester, England: Wiley. <http://dx.doi.org/10.1002/9780470713181.ch8>
- Sperandio, J. C. 1971. Variation of operator's strategies and regulating effects on workload, *Ergonomics* 14(5): 571–577. <http://dx.doi.org/10.1080/00140137108931277>
- Stanton, N. A.; Salmon, P. M.; Rafferty, L. A.; Walker, G. H.; Baber, C.; Jenkins, D. P. 2005. *Human factors methods: a practical guide for engineering and design*. Surrey, England: Ashgate Publishing, Ltd.
- Stewart, S.; Holmes A.; Jackson P.; Abboud R. 2006, October. An integrated system for managing fatigue risk within a low cost carrier, in *Proceedings of the 59th Annual International Air Safety Seminar (IASS)*, 23-26 October, Paris, France.
- Stokes, A.; Kite, K. 1994. *Flight Stress: Stress, Fatigue, and Performance in Aviation*. Aldershot, England: Ashgate, 50–95.
- Strayer, D. L.; Kramer, A. F. 1990. Attentional requirements of automatic and controlled processing, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(1): 67–82. <http://dx.doi.org/10.1037/0278-7393.16.1.67>
- Thackray, R. I. 1981. The stress of boredom and monotony: a consideration of the evidence, *Psychosomatic Medicine* 43(2): 165–176.
- Van Orden, K. F.; Limbert, W.; Makeig, S.; Jung, T. P. 2001. Eye activity correlates of workload during a visuospatial memory task, *Human Factors* 43(1): 111–121. doi: <http://dx.doi.org/10.1518/001872001775992570>
- Van Roon, A. M.; Mulder, L. J.; Althaus, M.; Mulder, G. 2004. Introducing a baroreflex model for studying cardiovascular effects of mental workload, *Psychophysiology* 41(6): 961–981. <http://dx.doi.org/10.1111/j.1469-8986.2004.00251.x>
- Vicente, K. J.; Thornton, D. C.; Moray, N. 1987. Spectral analysis of sinus arrhythmia: A measure of mental effort, *Human Factors* 29(2): 171–182. <http://dx.doi.org/10.1177/001872088702900205>
- Vidulich, M. A. 2002. Mental workload and situation awareness: essential concepts for aviation psychology practice, in P. Tsang, M. Vidulich (Eds.). *Principles and practices of aviation psychology*. Mahwah, NJ: Erlbaum, 115–146.
- Weltman, G.; Smith, J. E.; Egstrom, G. H. 1971. Perceptual narrowing during simulated pressure-chamber exposure, *Human Factors* 13(2): 99–107. <http://dx.doi.org/10.1177/001872087101300202>
- Wickens, C. D. 1980. The structure of attentional resources, in R. Nickerson (Ed.). *Attention and performance VIII*. Hillsdale, NJ: Erlbaum, 239–257.
- Wickens, C. D. 1984. Processing resources in attention, in R. Parasuraman, R. Davies (Eds.). *Varieties of attention*. New York, NY: Academic Press, 63–101
- Wickens, C. D. 1991. Processing resources and attention, in D. L. Damos (Ed.). *Multiple-task performance*. Boca Raton, FL: CRC Press, 3–34.
- Wickens, C. D. 2002. Multiple resources and performance prediction, *Theoretical issues in Ergonomics Science* 3(2): 159–177. <http://dx.doi.org/10.1080/14639220210123806>
- Wickens, C. D. 2008. Multiple resources and mental workload, *Human Factors* 50(3): 449–455. <http://dx.doi.org/10.1518/001872008X288394>
- Wickens, C. D.; Colcombe, A. 2007. Dual-task performance consequences of imperfect alerting associated with a cockpit display of traffic information, *Human Factors* 49(5): 839–850. <http://dx.doi.org/10.1518/001872007X230217>
- Wickens, C. D.; Goh, J.; Helleberg, J.; Horrey, W. J.; Talleur, D. A. 2003. Attentional models of multitask pilot performance using advanced display technology, *Human Factors* 45(3): 360–380. <http://dx.doi.org/10.1518/hfes.45.3.360.27250>
- Wickens, C. D.; Gordon, S. E.; Liu, Y. 2004. Stress and Workload, in C. D. Wickens, S. E. Gordon, Y. Liu (Eds.) *An introduction to human factors engineering*. New York, NY: Longman, 324–350.
- Wickens, C. D.; Hyman, F.; Dellinger, J.; Taylor, H.; Meador, M. 1986. The Sternberg memory search task as an index of pilot workload, *Ergonomics* 29(11): 1371–1383. <http://dx.doi.org/10.1080/00140138608967252>
- Wickens, C. D.; Liu, Y. 1988. Codes and modalities in multiple resources: A success and a qualification, *Human Factors* 30(5): 599–616. <http://dx.doi.org/10.1177/001872088803000505>
- Wickens, C. D.; Stokes, A. F.; Barnett, B.; Hyman, F. 1988. Stress and pilot judgment: An empirical study using MIDIS, a

- microcomputer-based simulation, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 32(2): 173–177. <http://dx.doi.org/10.1177/154193128803200238>
- Wiener, E. L.; Curry, R. E. 1980. Flight deck automation: Promises and problems, *Ergonomics* 23(10): 995–1011. <http://dx.doi.org/10.1080/00140138008924809>
- Wilkinson, R. T. 1964. Artificial signals as an aid to an inspection task, *Ergonomics* 7: 63–72. <http://dx.doi.org/10.1080/00140136408930725>
- Wilson, G. F.; Eggemeier, F. T. 1991. Psychophysiological assessment of workload in multi-task environments, in D. L. Damos (Ed.). *Multiple-task performance*. Boca Raton, FL: CRC Press, 329–360.
- Wilson, G. F.; Russell, C. A. 2007. Performance enhancement in an uninhabited air vehicle task using psychophysiologicaly determined adaptive aiding, *Human Factors* 49(6): 1005–1018.
- Wilson, J. R.; Rutherford, A. 1989. Mental models: Theory and application in human factors, *Human Factors* 31(6): 617–634. <http://dx.doi.org/10.1177/001872088903100601>
- Woods, D. D. 1996. Decomposing automation: Apparent simplicity, real complexity, in R. Parasuraman, M. Mouloua (Eds.). *Automation and human performance: Theory and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, 3–17.
- Yerkes, R. M.; Dodson, J. D. 1908. The relation of strength of stimulus to rapidity of habit-formation, *Journal of comparative neurology and psychology* 18(5): 459–482. <http://dx.doi.org/10.1002/cne.920180503>