

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~



# Defense Intelligence Reference Document

*Defense Futures*

15 December 2010

ICOD: 8 September 2010

DIA-08-1101-001

## Cognitive Limits on Simultaneous Control of Multiple Unmanned Spacecraft

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

## Cognitive Limits on Simultaneous Control of Multiple Unmanned Spacecraft

The **Defense Intelligence Reference Document** provides non-substantive but authoritative reference information related to intelligence topics or methodologies.

**Prepared by:**

(b)(3):10 USC 424

**Defense Intelligence Agency**

**Author:**

(b)(6)

COPYRIGHT WARNING: Further dissemination of the photographs in this publication is not authorized.

This product is one of a series of advanced technology reports produced in FY 2010 under the Defense Intelligence Agency, (b)(3):10 USC 424 Advanced Aerospace Weapons System Applications (AAWSA) Program. Comments or questions pertaining to this document should be addressed to (b)(3):10 USC 424;(b)(6) AAWSA Program Manager, Defense Intelligence Agency, ATTN: (b)(3):10 USC 424 Bldg 6000, Washington D.C. 20340-5100

## Contents

Summary.....	iv
Chapter 1: Introduction.....	1
Chapter 2: Measurement of Mental Workload.....	3
Subjective Measurements .....	3
Performance Measures.....	4
Physiological Measures .....	4
Cardiac Function .....	5
CNS Measurements .....	6
Ocular Measurements.....	6
Skin Measurements .....	7
Serum Levels of Hormones.....	7
Chapter 3: Studies in Cognitive Workload for Air Traffic Controllers .....	8
Modeling the Air Traffic Control Task .....	15
Chapter 4: Studies in Command of Multiple Semi-Automated Vehicles.....	18
Chapter 5: Discussion.....	22
Chapter 6: Conclusions.....	23
References .....	24

## Figures

Figure 1. One-dimensional Representation of Changes in Performance as Workload Varies.....	2
Figure 2. Typical EKG Signal for a Normal Heartbeat.....	5
Figure 3. Air Traffic Control .....	10
Figure 4. Representation of Performance Results from Brookings Study.....	13
Figure 5. Information Processing Model for a Human Operator.....	16
Figure 6. Examples of Unmanned Military Vehicles .....	19

## Tables

Table 1. Variables used in Determining Complexity of the Traffic. ....	11
Table 2. Correlations among Physiological Variables .....	14

## **Cognitive Limits on Simultaneous Control of Multiple Unmanned Spacecraft**

### **Summary**

Space exploration 40 years into the future may include manned missions to parts of the outer solar system. A possible scenario may include sending a small fleet of craft with different primary missions. For example, a trailing spacecraft of nuclear powered electromagnets designed to shield the manned part of the fleet from solar radiation; halo spacecraft with powerful radars to scout for incoming objects; exploration and mining craft, etc. The fleet could regularly travel out of unaided visual range of each other, joining up when necessary for maintenance, exchange of materials such as fuel, or other necessities. Piloting these multiple craft could be economically accomplished if only one remote pilot on station at a time was necessary.

The cognitive limitation of a human astronaut and his ability to perform the multiple-vehicle piloting task is the focus of this paper as little work has been done in this specific area. However, a large body of cognitive research on the limitations of object supervision and tracking for the task of air traffic control (ATC) exists. Additionally, there is an emerging body of research concerned with multiple unmanned vehicle piloting for heterogeneous missions. These are the two areas reviewed in detail as they relate to possible spacecraft missions.

Pilots develop an internal mental representation of the identity, position, mission, and current direction of relevant objects. This is referred to colloquially as "the big picture." The primary research question we seek to answer is whether there is a cognitive limit to the number of objects that can be monitored and tracked within the big picture. Secondly, we seek to find whether this maximum number is limited by the complexity of interaction; how those limiting factors are described, whether there is a real-time objective measure that indicates when a pilot is approaching his maximum capacity, and whether that capacity has been exceeded. The maximum number of tracked objects is highly dependent on the complexity of the piloting and mission tasks at hand.

Research is lacking in the area of cognitive limits on the number of spacecraft one pilot could control given any mission scenario. Currently, two models are being used to examine similar activities in air traffic control and remote piloting of multiple unmanned vehicles. In both areas it has been shown the cognitive limits on the number of craft capable of simultaneous control is 16 for simple destination selection, 7 for moderately complex piloting and/or mission task completion, and 4 for complex heterogeneous craft. While additional future research may help to increase the automation component of aircraft and mission control, no current evidence exists to show that a complete mental picture can be maintained for more than about 16 objects at one time, even with external working memory augmentation. However, it has also been demonstrated that physiological variables can be objectively employed to indicate overload. Nominal success has been achieved in classifying physiological states near high workload thus enabling both prediction and possibly prevention of overload.

## Chapter 1: Introduction

Due to the complexity, duration and numerous support requirements of future manned deep-space missions involving exploration, mineral exploitation, and possible colonization, a likely scenario will be the inclusion of unmanned fleets of support craft. Coupled with other requirements, an intensive research program is needed to investigate the cognitive limits on pilots and other operators responsible for the simultaneous control of multiple unmanned spacecraft making up the support fleet; a "fleet" approach is proposed in an effort to optimize safety and exploratory reach. This research effort would also aim at maximizing the functional efficiency of the mission and reducing the operation costs of unmanned vehicle fleets.

In this scenario there is much about the ancillary craft that are automated in both navigation and mission. Many of them will not require full-time piloting but given that they could be hundreds of miles from each other at any instant of time, they need monitoring to prevent unseen system failure or collision from letting them just disappear one day during the mission like a Martian probe. Accomplishing this monitoring task and the occasional piloting task for multiple craft in the fleet could be economically accomplished if only one remote pilot on station at a time was necessary.

We will focus here on the cognitive limitation of a human astronaut to perform the multiple-vehicle piloting task. It is not a surprise that there is little work in this specific area – in fact there were zero peer-reviewed articles in the major journals concerning remote piloting of multiple spacecraft (published in the last 30 years). There is however, a large body of cognitive research on the limitations of object supervision and tracking for the task of air traffic control (ATC). There is additionally an emerging body of research concerned with multiple unmanned vehicle piloting for heterogeneous missions. These are the two areas reviewed in detail as they relate to possible spacecraft missions.

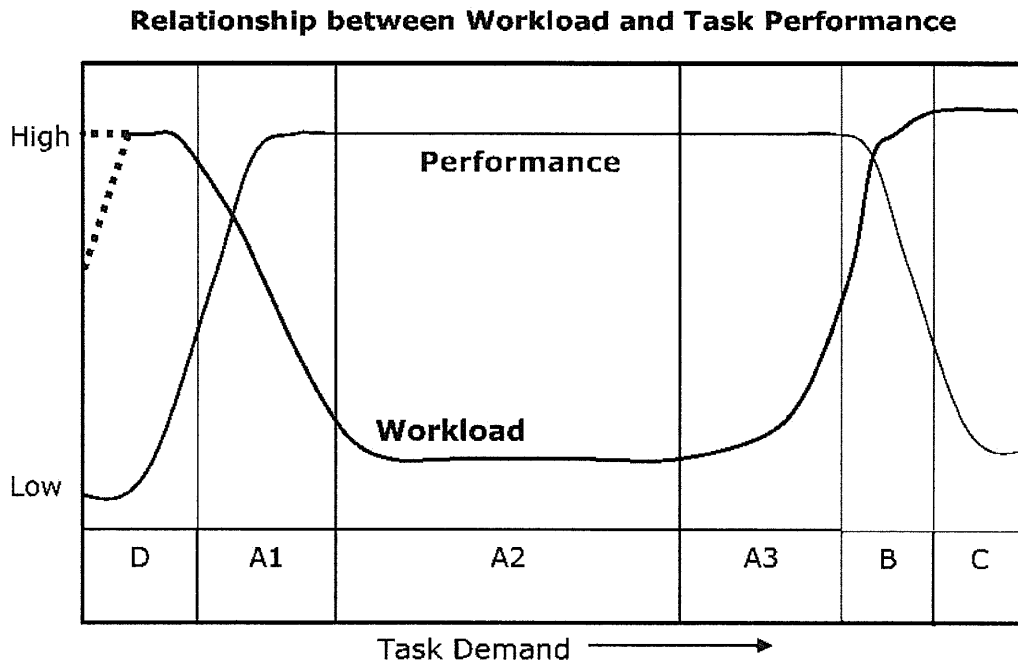
When a pilot or ATC operator is in control of several craft they have developed an internal mental representation of the identity, position, mission, and current direction of each object tracked. This is referred to colloquially as "the big picture." This mental representation is also called situational awareness. Keeping all of the information about all of the objects straight as long as they are in scope is the goal.

The primary research question we seek to answer is whether there is a cognitive limit to the number of moving objects that can be maintained in the big picture. Secondary questions are whether this maximum number is limited by complexity, how those limitations might be described, whether there is a real-time objective measure that will indicate when a pilot is approaching their maximum capacity, and whether that capacity has been overloaded.

For the current treatise we will consider only traditional humans as pilots. Cyborg-enhanced astrobots are a topic for another tome.

In discussing cognitive limitations, it is useful to introduce the concepts of task demand, mental workload, and a simplistic model of multiple resource theory. For a given task, the gross level of neural activity required is a representation of the mental demands of the task. Of course, a complex task can demand varied resources such as visual and

audio processing, and this is where multiple resource theory enters: different resources can be considered independent if demands on one resource do not tax the availability of capacity of the other. Performance of a task can be high or low: in general if spare capacity is available, performance is high, and if capacity is limited or exceeded, performance is low. Moreover, task demand enforces a fixed theoretical relationship between mental workload and task performance: this is shown in Figure 1.



**Figure 1. One-dimensional Representation of Changes in Performance as Workload Varies.** Task demand increases to the right. In the three A regions, performance remains unchanged: A2 is optimal, where a trained operator exerts minimal effort to maintain a set level of performance. At times of low demand the operator exerts effort to maintain vigilance (A1) until demand is so low, the subject disengages from the task (D). The workload curve is not well defined at very low levels of demand (dotted lines). At the other end, as demand increases, the subject can exert effort to keep up with demand (A3). Additional effort maintains performance until degradation begins (B), and in the overload condition, region C, performance is degraded beyond acceptable levels. While effort is maintained in overload, some low level of performance exists.

## Chapter 2: Measurement of Mental Workload

There are three approaches to measuring mental workload in a subject performing a primary task. The first is subjective evaluation, either post-hoc self-report questionnaires concentrating on how "busy" one may have felt, or an experimental observation of activity. The second is performance measures, an objective evaluation of how well the subject completes the primary task, a secondary task, or an experimentally inserted reference task. The final approach is to record real-time physiological measures, with the assumption that increased workload increases anxiety and this will be exhibited by changes in the autonomic nervous system (ANS).

### SUBJECTIVE MEASUREMENTS

Self-report measures are appealing because they get inside the mind that was performing the task. There is no absolute objective scale to measure one person's "fully occupied" from another person's view of the same state; however, through rating scales and self-drawn graphs, one can obtain an accurate picture of how perceived workload evolved during the experiment. Perceived workload is important because it is what needs to be maintained between a subjective minimum, where attention may wander, and a subjective maximum, where increased emotion may decrease performance capacity.

The most frequently used standardized self-report tools are the NASA Task Load Index (NASA-TLX or just TLX) and the Subjective Workload Assessment Technique (SWAT).<sup>1,2,3,4</sup> The TLX is a subjective workload assessment based on a multi-dimensional rating questionnaire. An overall workload score is derived based on a weighted average of ratings on six subscales: mental demands, physical demands, temporal demands, own performance, effort, and frustration. SWAT is a two-step assessment of three workload factors: time load, mental effort load, and psychological stress load. In the first step, hypothetical activities are ranked according to perceived workload. In the second step, the experimental task is evaluated post-hoc, using a 1-3 rating scale for each of the three dimensions. An interval scale of workload is derived, from 0-100, based on the reference data collected for each subject in the first step, and the evaluation of the experimental task. A custom self-report can also be designed by the experimenter to specifically focus on the research questions in a given experiment.

The second type of subjective measure is evaluation by an expert observer. In this type of measurement, assumptions are made on the mental activity of the subject based on the activities being performed. The advantage of this approach is that there is minimal variance per-to-person if the same evaluator is employed. The disadvantage of this approach is the outside observer can miss workload multiplying factors such as task complexity contributing to an overwhelmed feeling by the subject. An example of this approach is the concept of *utilization*. In calculating utilization it is assumed that the subject must cognitively address one issue at a time in serial order. The measure of utilization is percent time busy, or addressing any issue, as opposed to waiting or monitoring for the next event. In general it is observed for control and supervisory tasks that at around 70% utilization performance begins to degrade.<sup>5</sup> Arguably not a perfect measure of workload, utilization has the advantage of simplicity, objectivity,

and a quantitative scale, allowing it to be used in threshold detection and prediction, as well as in computer-assisted workload balancing.

## **PERFORMANCE MEASURES**

Laboratory, reaction time to an event stimulus is a typical real-time measure of speed. Outside the laboratory, in a more natural environment, speed is an indirect measure of a subject's ability to keep up with a given rate of events.<sup>a</sup> Accuracy is measured in both the laboratory and naturalistic environments: for example, in an air-traffic control task, properly handing off a plane to the next controller, as it leaves the first controller's monitored airspace<sup>b</sup>, is measured as a successful task completion.

Performance, when utilizing a secondary task, is accomplished following two paradigms. First, in the dual-task paradigm, performance on the secondary task is required and primary task performance is thus an indication of workload. For the second paradigm, instruction is given to maintain the primary task performance, and performance on the secondary task is thus a measure of "space capacity" for additional workload.

Care is required in selection of secondary tasks in order to ensure they affect the resources one wishes to probe. For example, in a driving scenario, a minimally intrusive task is to push a button on the floor or steering wheel when a light flashes in the field-of-view of the driver. This visual "detect-and-respond" task adds to what is primarily a complex visual and motor coordination task that probes the capacity of visual attention.<sup>6</sup> This is different than a secondary task utilizing different resources, such as having a conversation while driving. When planning the paradigm, a model needs to be developed that treats (or explicitly ignores) individual tasks and interactions among them.

Reference tasks are executed before and after primary tasks. Typically reference tasks focus on trending in performance due to effects such as fatigue. One important case of reference tasks is to normalize an individual's current capacity for mental workload. Such a pre-performance measure can be used to adjust maximum workload to compensate for day-to-day variation.

## **PHYSIOLOGICAL MEASURES**

The human nervous system is anatomically divided into the Central Nervous System (CNS) – and the Peripheral Nervous System (PNS). The CNS includes the brain and the spinal cord. The PNS is made up of the somatic division, which innervates the skin, voluntary muscle, and joints, and the autonomic division, which mediates visceral sensation as well as executes motor control of smooth muscle, viscera, and endocrine glands. The autonomic division consists of sympathetic, parasympathetic, and enteric systems. The sympathetic system mediates response to stress, while the parasympathetic system works to maintain homeostasis and conserve body resources.

<sup>a</sup> Reaction time is not typically measured in real-time in the field as events occur at unplanned times. Post-hoc analysis can resolve reaction times down to a comparable resolution to innate motor reaction variance, on the order of 10's of msec.

<sup>b</sup> Contrary to what a layperson may think, the vast majority of errors in air traffic control do not result in collisions or even near misses; rather they are mistakes in procedure.

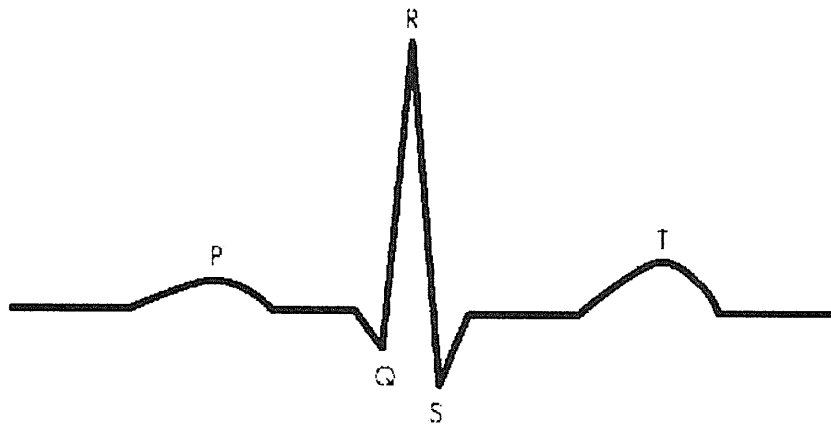


The enteric system executes control of smooth muscle. Although the CNS and PNS are anatomically distinct, they are functionally intertwined. When discussing the function of the autonomic division, it is customary to refer to it as the Autonomic Nervous System, or ANS.<sup>7</sup>

Changes in global arousal or activation through changing workload can result in changes in physiological activity. These measures are advantageous as changes are measurable continuously, in real-time, and usually unobtrusively in a naturalistic setting. The drawback of using physiology alone is that there is no direct measure of primary task performance.

### Cardiac Function

Normal beating of the human heart produces a distinct and repeating pattern of electrical activity measurable throughout the body for any two sample points that cross the chest.<sup>c</sup> The typical electrocardiogram signal is shown in Figure 2.



**Figure 2. Typical EKG Signal for a Normal Heartbeat.** Portions of the waveform are labeled P, Q, R, S, and T.

Detection of the R-wave allows measurement of frequency, time,<sup>d</sup> and amplitude. For continuous monitoring, heart rate measurements will vary considerably and in a non-linear fashion; therefore, the measurement of inter-beat-interval (IBI), the time between R peaks, is more normally distributed in the absence of signal, reducing noise in the measurement.<sup>8</sup> Averaging the heart rate over minutes of task performance and comparing to baseline yields a reliable estimate of increased metabolic function.<sup>9</sup> An additional measure is the heart rate variability (HRV), calculated by dividing the standard deviation of IBI by an average value of IBI within a sample period. Additional measurements can be made by decomposing the spectra of HRV into low, mid, and

<sup>c</sup> In fact, a typical introductory physics course for life science students will include a laboratory experiment where the heart rhythm is measured between electrodes located on the right wrist and left ankle.

<sup>d</sup> This time measurement is properly a phase measurement, with the phase relative to some reference event.

high frequency components which have different noise contributions (for example, core temperature changes, blood pressure or speech, and respiration, respectively).<sup>10</sup>

Blood pressure and its variability can also be measured. For continuous monitoring, a finger cuff filled with water match the inner-arterial pressure and can be used to monitor variability.<sup>11</sup>

## **CNS Measurements**

Measurement of brain activity can be unobtrusively recorded using low-count electroencephalography (EEG),<sup>e</sup> or the minimally obtrusive techniques of EEG, near-infrared spectrometry (NIRS), trans-cranial Doppler sonography (TCDS), or the non-invasive laboratory techniques of functional MRI (fMRI), magnetoencephalography (MEG), or positron emission tomography (PET). The latter three techniques are for brain research only and do not have any current naturalistic research studies (although see Genik<sup>12</sup> for a prognosis on generation-after-next technologies including NIRS, PET, fMRI, and MEG).

EEG measurements are typically divided into spectra and the relative power in the bands 0-4 Hz ( $\Delta$ ), 4-8 Hz ( $\theta$ ), 8-13 Hz ( $\alpha$ ), 13-30 Hz ( $\beta$ ), and 30-100 Hz ( $\gamma$ ). NIRS measures the BOLD effect<sup>f</sup> and is related to localized  $\gamma$  activity. TCDS measures CO<sub>2</sub> as the byproduct of localized increased metabolism and is also an indirect measure of neural activity which has been shown to be related to vigilance.<sup>13</sup> For EEG experiments in mental workload, changes are typically reported in the  $\theta$  and  $\alpha$  bands, though more recently  $\beta$  and  $\gamma$  bands have shown sensitivity.<sup>14</sup>

Event-related potentials (ERP) are peaks of activity measured at the skin indicative of several tens of thousands of neurons firing coherently for a short time. For example, a well-studied cognitive ERP is P300, a 10's-of-msec-wide bump in the EEG signal occurring 200-400 msec after an event. Some success in utilizing a task-irrelevant secondary audio stimulation and N100 has been shown,<sup>15</sup> but little further development to this approach has been found in the last 15 years.

## **Ocular Measurements**

Measurements involving eye fixations, dwell time (temporal length of a fixation), and pupillary changes are well established metrics of workload in visual searching tasks.<sup>16</sup> Additional measures of ocular changes include blink rate, blink duration, blink latency, and eye movement. These are recorded using one of various types of eye-tracking (ET) or electrodes to measure an electrooculogram (EOG). ET data will include position of a fixation and the time of each eye movement (or saccade), whereas an EOG only identifies the time that the muscle controlling eye blinks or eye position activated.

---

<sup>e</sup> Low-count EEG is generally less than 10 electrodes, and usually 3 or 5.

<sup>f</sup> The Blood Oxygen Level Dependent (BOLD) effect is a local change in the oxygen saturation ratio near neural activity due to metabolic and vascular action. This change is detectible in the infrared spectra.

### **Skin Measurements**

The ANS controls the opening and closing of sweat glands in response to stress and anxiety. These changes can be observed by measuring electrodermal activity, typically skin conductance response (SCR),<sup>17</sup> but also skin potentials (SP) and skin temperature (ST). A standard score of these measures can be calculated by collecting resting state data prior to the task interval.

### **Serum Levels of Hormones**

Hormone levels are direct results of activity in the ANS. It is natural to want to continuously monitor certain stress hormones such as adrenaline or cortisol. The difficulty lies in measurement time – typical fast assays of salivary cortisol still require about 15 minutes.<sup>18</sup> Until reliable in-situ monitoring is available, serum levels of hormones will only play a supporting role in establishing baseline capacities or post-incident analysis.

### Chapter 3: Studies in Cognitive Workload for Air Traffic Controllers

There are a few areas of research applicable as analogues to the piloting of several aircraft. The largest of these areas analyzes the supervisory functions of air traffic controllers (ATC). In a typical ATC duty environment, a single operator is responsible for many fully-autonomous aircraft. ATC is either airport-based or en-route. Airport-based facilities include ground control, local control for active runways (Tower control), and approach control, which in the US is called Terminal Radar Approach Control, or TRACON. A zone of air-controlled space assigned to a specific control center is that center's *sector*; similarly, a given controller is responsible for a sector of airspace. Between airport-controlled sectors, aircraft are monitored by an en-route facility. TRACON is the most cognitively demanding function in this chain. Terminal controllers, as ATCs are called at TRACON facilities, are responsible for departures after take-off, and approaches and flyovers within about a 50-mile radius of the airport.<sup>9</sup> Approaching aircraft need to be vectored by the controller into an appropriate flight path for landing, avoiding all other aircraft in the air or soon to be in the air, and then handed off to the Tower controller for landing and ground instruction. Departing flights and flyover traffic mainly need to be monitored for conflicts. Approaching aircraft are by far the most cognitively challenging in the ATC task. Images of air traffic controller environment and displays are shown in Figure 3.<sup>h</sup>

Errors in flight control are called anomalies in the industry. The most common aircraft anomaly is deviation from flightpath in en-route sectors. This anomaly is corrected by pilots themselves or after instruction from the appropriate ATC. The most common ATC error is miscommunication between controllers in an aircraft handoff between sectors.<sup>19</sup> A study in 1997 by the National Academies showed that ATC errors occurred in both high and low workload conditions, as predicted by overload and disengagement.<sup>20</sup>

In looking at the cognitive limits in ATC, we should seek where typical controllers enter the B region of workload. Traffic load defined as simply the number of aircraft does not by itself show the complete picture of ATC workload. In a study of professional controllers in 2006, Boag recorded subjective measures of workload and reaction time when static air traffic displays were presented. Displays included air traffic conflicts of differing complexity that required resolution. Complexity in the display was objectively ranked using the Method of Analysis of Relational Complexity (MARC).<sup>21</sup> Results showed that a relatively small number of aircraft can greatly increase the perceived workload. Boag concludes that perceived complexity is the number one factor in determining workload, and that although conflicts are the major source of complexity in the ATC task, and these can be modeled somewhat using a combination of aircraft separation and transition<sup>i</sup> variables, individual differences still have significant impact on when a controller may reach overload.<sup>22</sup>

---

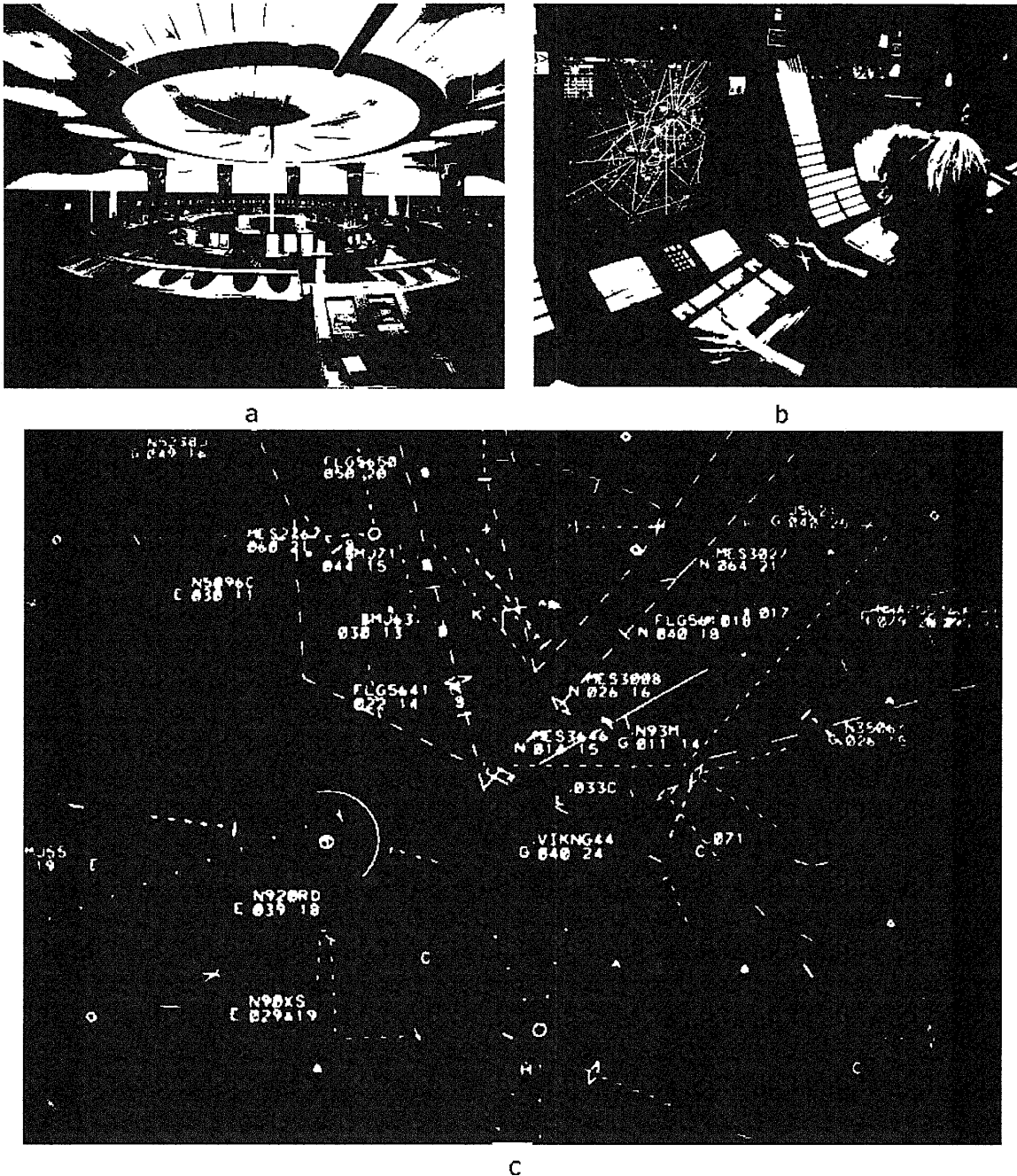
<sup>9</sup> Each facility will vary in TRACON sector radius. For smaller airports, TRACON functions may be performed by a nearby en-route facility.

<sup>h</sup> Of historical note in the US Air traffic Control industry was the industry-wide strike begun on August 3, 1981, by the nearly 13,000 ATC specialists. Only 1300 obeyed a Presidential order to return to work under the "peril to national safety" provision of the 1947 Taft-Hartley Act. On August 5, 1981 the remaining 11,345 controllers were fired and banned for life from federal service. The FAA rebuilt the force to pre-strike levels during the rest of the 1980s. This event resulted in a dearth of research during the 1980s on air traffic control professionals.

<sup>i</sup> An aircraft transition is an event such as landing, hand-off to another controller, or entering/leaving a sector, where a sector is defined as a controller's airspace of responsibility.

In a post-hoc examination of sources of human error in taped controller data, Chang developed a conceptual model of ATC task within a system of human, software, hardware, and environment using the SHEL<sup>23</sup> approach of ergonomics. This study also examined several aspects of personnel management, but when overload was examined as the source of error, it was shown that all factors need to be considered along with their interaction. That is, the human and the external task cannot be considered in isolation of the environment of the other humans, and the augmentation capabilities of software, hardware, and the performance environment significantly affect error rates.<sup>24</sup>

Operator overload can result in subsequent errors if sufficient recovery time is not allotted. Di Nocera examined specific error rates after short periods of overload. The first part of the study successfully proved that these so-called post-completion errors existed in times immediately after peaks in workload. The second part of the study examined whether an augmentation tool to assist in medium term conflict detection could reduce these errors. The subject population included 18 military ATC of varying experience. Workload was based on self-reported NASA-TLX. Results showed that augmentation processes helped junior controllers, but senior controllers were unaffected by the assistance.<sup>25</sup>



**Figure 3. Air Traffic Control:** a) overview of a large TRACON facility. Potomac is pictured. b) A single ATC TRACON station. Beside the radar display is a stack of memory aid blocks when the controller writes essential aircraft information – as long as there is a physical block to the right, there should be a corresponding “blip” on the screen. c) TRACON display from Minneapolis-St. Paul. On this display are aircraft under control (undesignated), as well as craft from 3 other controllers, East (E), North (N), and Ground (G). Below each aircraft name is controller designation, altitude in hundreds of feet, and airspeed in tens of knots.

Lamoureaux researched in 1999 whether a proposed ATC augmentation system could safely allow smaller distances between aircraft and thus increase traffic capacity without building new airports or runways. Pairs of aircraft were characterized based on their direction of travel and separation as shown in Table 1. During the task, operators rated their instantaneous perceived workload on a five-point scale roughly corresponding to D, A1, A2, A3/B, and B/C.<sup>j</sup> Operators were prompted for their self-assessment once every two minutes. The experiment was able to generate self-assessment values between 1 and 4; no overload conditions were generated in this study. Using a model based on task variance and complexity of current aircraft under control, the researchers were able to predict perceived workload 74% of the time, but more importantly, were able to predict a self-rating of 4 for 80% of the cases. Predicting the boredom value of 1 was less successful, at about 60%. Besides the numerical predictions, the authors conclude that complexity of the task drives perceived workload.<sup>26</sup>

**Table 1. Variables used in Determining Complexity of the Traffic** relationship between pairs of aircraft. Using four variables with three thresholds gives different classes of complexity. For example, an aircraft pair could be 5 miles apart (3 to 7 threshold), traveling in the same lateral direction, flying with 2500 ft of vertical separation (> 2000ft), and both straight and level. There are 81 such combinations.

Variable	Threshold	Low	Mid	High
Lateral separation		< 3 miles	3 to 7 miles	> 7 miles
Lateral direction		Same direction	Opposite direction	Crossing
Vertical separation		< 800 ft	800 to 2000 ft	> 2000 ft
Vertical direction		Both straight and level	One straight, one climbing or descending	Both climbing or descending

Physiological indicators of stress were measured at two low-traffic control centers (Fayetteville, AR and Roswell, NM) and one higher traffic center (Oklahoma City). Heart rates were measures along with hormone secretion levels in urine. The urine specimens were pooled before analysis into two groups throughout a 5-day workweek: during the 8-hour workday, night time after work. Additionally, controllers completed the State-Trait Anxiety Index before and after each workday. Results showed that lower traffic centers exhibited lower stress levels, and that the best biological indicator of stress was epinephrine levels from urine rather than HR. The authors concluded that traffic load and complexity were the main sources of stress in the ATC task rather than the nature of the job itself.<sup>27</sup>

A straightforward experiment to look at physiological responses using EEG and EOG to lapses in attention during ATC tasks was performed by Peiris in 2005. The goal of the study was to categorize expert analysis of EEG/EOG data to develop an automated analysis program that could read the data online without human intervention to alert operators to attention lapses and low levels of alertness. Professional ATC operators were given 10 minute intervals of the psychomotor vigilance task (PVT). Recorded data was evaluated by several human expert EEG and EOG analyzers. These experts were not able to correctly identify alertness or attention lapses (EEG identified only 6 of 101

<sup>j</sup> In this study, the A3/B rating described "Non-essential tasks suffering, could not work at this level for long," while B/C described getting behind and losing situational awareness.

lapses). Peiris concluded that a built-from-scratch automated system is needed to identify subtle features, especially in the low-count electrode EEG (5 electrodes).<sup>28</sup> In the definition of complexity, it is important to note that one should not focus entirely on a single aspect of the ATC task in the laboratory. Donald defines complexity in two aspects: task complexity independent of the event rate, and ancillary aspects of the job in a naturalistic environment. The complexity of the task as a whole needs to be considered instead of focusing on a single aspect such as monitor and detect and the rate at which this task can be completed.<sup>29</sup>

The naturalistic environment was in fact utilized by Brookings in 1996.<sup>30</sup> Moreover, *in situ* measurements were conducted by Collet in 2009.<sup>31</sup> Both studies examined TLX ratings, as well as several ANS variables. Brookings additionally utilized EEG.

In the Brookings study, three simulated TRACON sessions were conducted. The first session varied traffic volume between low, medium and high levels, while the second session varied task complexity at a constant rate of aircraft. The third scenario was conducted with an overwhelming number of aircraft, the goal being to take physiological data in the condition where situational awareness is lost.<sup>k</sup>

The traffic load variance session lasted 45 minutes with three 15-minute sessions where the controller was required to handle 6, 12, and 18 aircraft; the order of presentation was counterbalanced across subjects. Other complexity factors, such as the ratio of overflights, arrivals, and departures, were kept constant.

In the complexity variation session, the number of aircraft was kept constant at 12, while various complicating factors were modulated. Changing complexity factors included:

- Altering the ratio of arriving to departing and flyover traffic.
- Changing the probability that a pilot didn't hear or failed to execute a controller's instruction.
- Increasing or decreasing the heterogeneity of aircraft type.

In the overload session, 15 aircraft were presented in 5 minutes.

Physiological variables monitored included heart activity using two electrodes on the chest, EOG using electrodes around the eyes,<sup>l</sup> respiration using elastic transducer bands, and 19 channels of EEG using a cap outfitted with a standard 10-20 configuration.<sup>m</sup> Task performance points were awarded for successfully handing aircraft, minus any points for operational errors such as separation conflicts, hand-off errors, and missed approaches. TLX ratings were recorded between workload conditions during a designed 1-minute lull in traffic. The simulation was considered quite difficult, even for professional Air Force ATC, and participants were required to practice until they didn't

<sup>k</sup> In colloquial terms, ATC call this "losing the picture."

<sup>l</sup> Electrodes are pointed out here as more recent methodology could use infra-red optical devices to record heart and ocular activity.

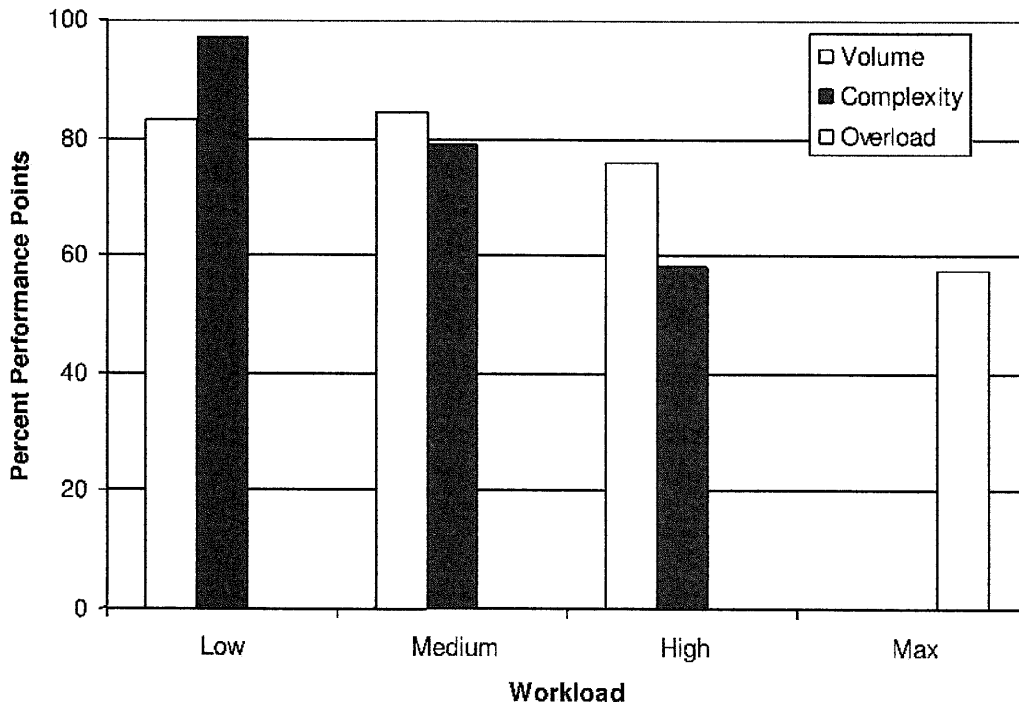
<sup>m</sup> The standard 10-20 configuration refers to electrodes every 10%/20% of the total distance between right-left/anterior-posterior anatomical markers. A 10-10 configuration would include twice the electrodes, etc.



crash any planes in any of the scenarios – this required approximately 6 hours of practice per participant before the experiment was conducted.

Only one of the eight controllers in the Brookings study rated the overload condition as a loss of situational awareness. The results compared the TLX, primary task performance, and physiological measures to low, medium, and high workload conditions, as well as the max or overload condition. Primary task performance is represented in Figure 4 (this chart was recreated visually from the source chart to accurately represent all trends), showing a trending effect for complexity but not volume. Additional results showed that changes in task difficulty (volume or complexity) produced changes in TLX, eye blink rate, respiration rate, and the EEG power spectra. The EEG power spectra were different for changes in volume versus changes in complexity. There were no observed significant correlations with heart rate or heart rate variability. The authors conclude that psychophysiological data can be used to accurately measure workload in real-time, an observation they note confirms earlier work on F4 crew members performing flight tasks of modulated complexity.<sup>32</sup>

The Brookings data was reanalyzed by Wilson in 2003 using an artificial neural network approach as well as a stepwise discriminate analysis to classify a physiological state as either operational or overloaded. Wilson successfully classified the overload condition consistently in more than 98% of the cases. The authors admit an issue with psychophysiological variation (day-to-day) that would need to be normalized and further research is required.<sup>33</sup>



**Figure 4. Representation of Performance Results from Brookings Study.** <sup>30</sup> Shown are the three scenarios with modulation of traffic volume, complexity, and the overload condition. Note that the Low workload entry for the volume modulation performance (6 planes) was already 80% and this was similar to the 12-plane medium complexity performance. Only the low complexity, 12-plane scenario showed near 100% primary task performance.

The more recent Collet study recorded 5 ANS variables from 25 participants during real ATC operations. The population, mean age of 44, included only fully qualified operators, who were monitored for one hour during TRACON duty at Saint Exupéry International Airport (Lyon, France). Correlation analyses were performed with the number of aircraft the operator was currently controlling. No adjustment was made for task complexity; however, data were acquired between 6 and 9 PM local time to collect medium and high workload data. Each participant handled between 1 and 10 aircraft during the study. The results of the correlation analysis are shown in Table 2. The authors conclude that changing the number of aircraft for professional ATCs produced correlations in physiological measures for SC, SBF, and IHR.<sup>31</sup>

**Table 2. Correlations among Physiological Variables** in a study of air traffic controller workload modulation with variable number of aircraft. NA: number of aircraft; TLX: NASA self-report workload metric; Std SC: normalized skin conductance; Std SP: normalized skin potential; Std SBF: normalized capillary blood flow measured through the skin; Std ST: normalized skin temperature; Std IHR: normalized instantaneous heart rate. Bold values show significant correlation. SC, SBF, and IHR show significant correlation with changes in NA. Normalizations (standardizations) were performed against baseline data per subject to decrease inter-subject noise.<sup>31</sup>

	NA	TLX	Std SC	Std SP	Std SBF	Std ST
NA	1					
<b>TLX</b>	<b>.98</b>	1				
	p<.001					
<b>Std SC</b>	<b>.93</b>	<b>.89</b>	1			
	p=.002	p=.008				
Std SP	.77	.67	<b>.91</b>	1		
	NS	NS	p=.005			
<b>Std SBF</b>	<b>-.97</b>	<b>-.94</b>	<b>-.87</b>	-.75	1	
	p<.0001	p=.001	P=.02	NS		
Std ST	-.79	-.80	-.62	-.43	.82	1
	NS	NS	NS	NS	NS	
<b>Std IHR</b>	<b>.98</b>	<b>.95</b>	<b>.97</b>	<b>.85</b>	<b>-.93</b>	<b>-.88</b>
	p<.0001	p<.0001	p<.0001	p=.03	p=.002	p=.005

Adaptive automation (AA) is the rebalancing of workload between the computer and human. Low workload levels can be supplemented with usually routine tasks that will keep the operator attentive, while providing the subject with additional mission information. This "extra information" may not be critical, but it will keep the subject from disengaging from the overall task. The goal of AA is to maintain peak performance of the system, in the A1 to A3 regions. Kaber studied AA in terms of a simulated ATC task in 2005. Forty non-professional participants were monitored for primary and a probe secondary task performances. Results showed that primary task performance was greatest when AA was added to the system.<sup>34</sup>

## **MODELING THE AIR TRAFFIC CONTROL TASK**

Like any profession, ATC personnel experience day-to-day variation in performance, and there are natural variations between controllers. In order to study these differences, mentioned in most of the studies detailed above, a model needs to be built of the controller, the environment, and the task, with the goal of locating where the majority of changes may be occurring, and where any augmentation may be best suited to assist in performance.

Specific to the air traffic controllers, the major variation source found when studying large variations in performance was disruption of the circadian rhythm leading to a disequilibrium condition described as a biological instability. Fortunately, no fancy technology system was required to solve this particular problem, just proper human resource management to avoid frequent shift switching.<sup>35</sup>

Loft proposed that modeling the ATC task complexity and workload is insufficient to predict performance due to the overriding effect of operator decision strategy. ATC operators can select priorities, manage their own cognitive resources, and thus regulate their own performance. The primary relief for the ATC operator is handing off traffic to another local operator.<sup>36</sup> Our overall topic is concerned with a single pilot in a space environment, where no room full of colleagues exists to take up the slack; therefore, such group modeling techniques are outside the scope of the current treatise.

We do note that Loft develops excellent single-task descriptions of time pressure, conflict detection, conflict resolution, etc.

As shown multiple times in the preceding section, single-task processing time and intensity (difficulty or complexity) are the primary drivers of workload. Developing a model connecting time, intensity, and effort, Hendy shows how decision time connects a time-intensity-effort loop (Figure 5). Hendy contends decision time is the single variable dominant in workload. Within this loop model, increasing the event rate is akin to increasing task difficulty. The adaptation strategies are developed with training and experience, a possible explanation of the difference in junior and senior performance with augmentation aids.

Averty contends that ATC workload cannot be directly measured, but must be inferred from a quantifiable mixture of including objective and subjective measures. He breaks down the controller task into monitoring, vectoring, and conflict solving, and develops a refinement of the NASA-TLX called TLI. Averty's Traffic Load Index is based on number of aircraft, but each aircraft is given additional weight according to processing requirements on the controller, including both cognitive and emotional weight: for example, aircraft with path conflicts to resolve are given the highest weight, while isolated flyover traffic is given low weight. The authors conclude that TLI needs to include physiological inputs as well to fully model the task-controller interaction.

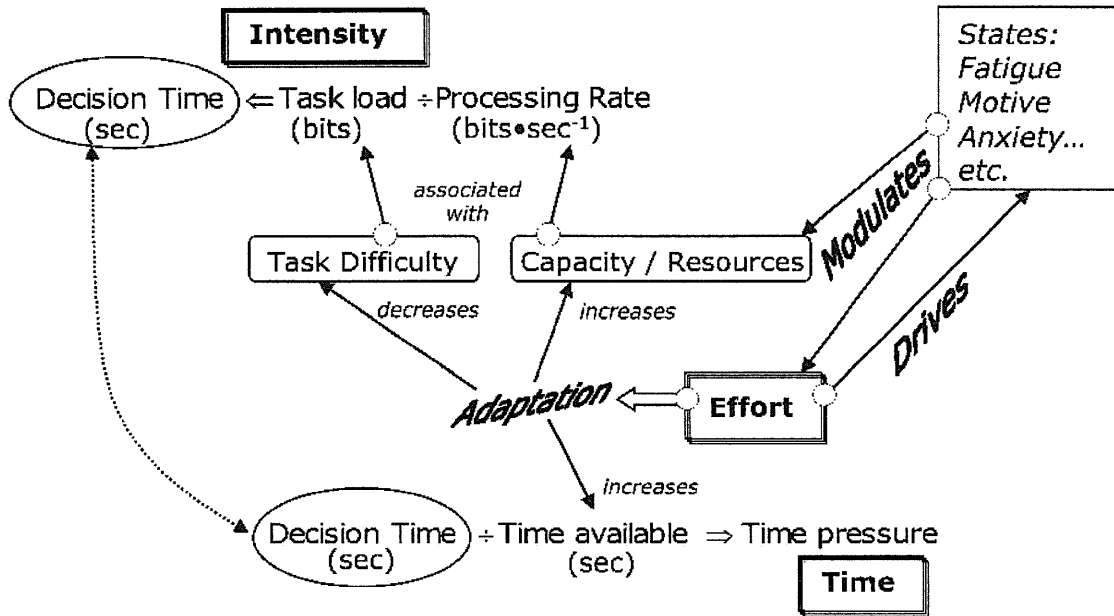


Figure 5. Information Processing Model for a Human Operator. <sup>37</sup>

In modeling the controllers themselves, one can look to selection efforts and find which candidate skills are the best predictors of training success. Pre-strike data (see footnote h on page 9) of ATC training showed that strong candidates had skills in spatial relations, abstract reasoning, and math as well as oral decision-making.<sup>38</sup>

In designing a model, it is tempting to engage subject matter experts (SME) to estimate the demand of various resources within a multiple resource model of ATC. Cohen studied SME predictions of workload within a model of 7 channels (visual perception, auditory perception, spatial information processing, analytical information processing, verbal information processing, manual activity and speech). The authors concluded that using the 7-channel multiple resource model with the SME approach doesn't work for predicting workload in the ATC task.<sup>39</sup>

Hancock has written extensively on the multiple-resource model of cognitive task performance under stress. He points out that functional brain imaging studies clearly show resources are separated anatomically, providing evidence that the multiple resource model should not be abandoned in future research.<sup>40</sup>

One model that was proposed was that local visual distractions interfered with cockpit ATC tasks. The model has attractiveness given the incredible complexity of cockpit displays and the popular notion that in-vehicle distraction is the root of all evil for automobile driving. Iona proposed that a tunnel display for in-cockpit ATC information would reduce the effect of outside visual distraction and increase primary task performance measures. The authors concluded that this type of augmentation has little effect on trained professional pilots in performance of their duties.<sup>41</sup>

In an essential foundation study for introducing forms of augmentation to the ATC task, Wickens modeled the dual task environment of pilot traffic avoidance using alarms to augment detection of conflicts. The primary task was maintaining aircraft flightpath using a simulated cockpit display of a crosshair inside a box: the crosshair indicated aircraft direction and drifted toward the sides of the box if not corrected by the pilot using a joystick control. The drift rate was variable and could increase or decrease the difficulty of the primary task. The computer monitored for potential collisions and warned pilots if another aircraft was within 3 miles.

Pilots were instructed to maintain their own aircraft flightpath first, and then detect conflicts. Upon alarm, the pilot was to examine an ATC display and recommend re-routing of the conflicting aircraft. Pilots were informed that the automated detection system may erroneously label some situations as conflicts; therefore the pilot needed to actually perform several ATC cognitive task functions to confirm the conflict before recommending action. Participants in the study were 12 student pilots.

The results of this experiment showed that when augmentation was correct more than about 80% of the time, performance decreased due to decreased vigilance in confirming alarms properly on the ATC display. Additionally, with a high accuracy in the alarm rate, pilots did not regularly check the ATC display to ensure that the augmentation didn't miss possible conflicts. Auditory and visual binary alarms were presented and the auditory alarms were more effective and did not interfere with the primary task (visual tracking). The authors concluded that a 20-25% false alarm rate was optimal<sup>n</sup> when it is intended for the pilot to work alongside the automation rather than rely on it.<sup>42</sup>

The Wickens study was undertaken with the plan of moving ATC to a shared responsibility of the controller and pilot: the goal being to increase airspace capacity by removing some of the more mundane functions like en-route course correction and en-route conflict detection to primarily cockpit control. This situation would be analogous to a spacecraft pilot operating their primary vehicle manually while attending many semi-automated ancillary vehicles. Although not exactly the same,

Landsdown studied TLX workload measures on drivers performing multiple in-vehicle tasks. The authors here concluded that secondary tasks significantly increased perceived workload in this arrangement of task control.<sup>43</sup>

---

<sup>n</sup> This noise in the signal is analogous to adjusting the squelch level on a CB radio: too high a setting and you miss traffic; too low a setting and all you hear is random noise.

## **Chapter 4: Studies in Command of Multiple Semi-Automated Vehicles**

Another analogue to remotely commanding several spacecraft is piloting or supervising operation of several unmanned ground or air vehicles (UGVs and UAVs). The primary use of these scenarios is in military operations which imposed additional criteria on the control system. Examples of several UAVs and their primary missions are shown in Figure 6. Control of vehicles can be either in-theater, at a range of yards to 10's of miles, or from a long-range command and control center, such as the Predator reconnaissance in the Iraq or Afghan Theater executed from bases within the continental United States.

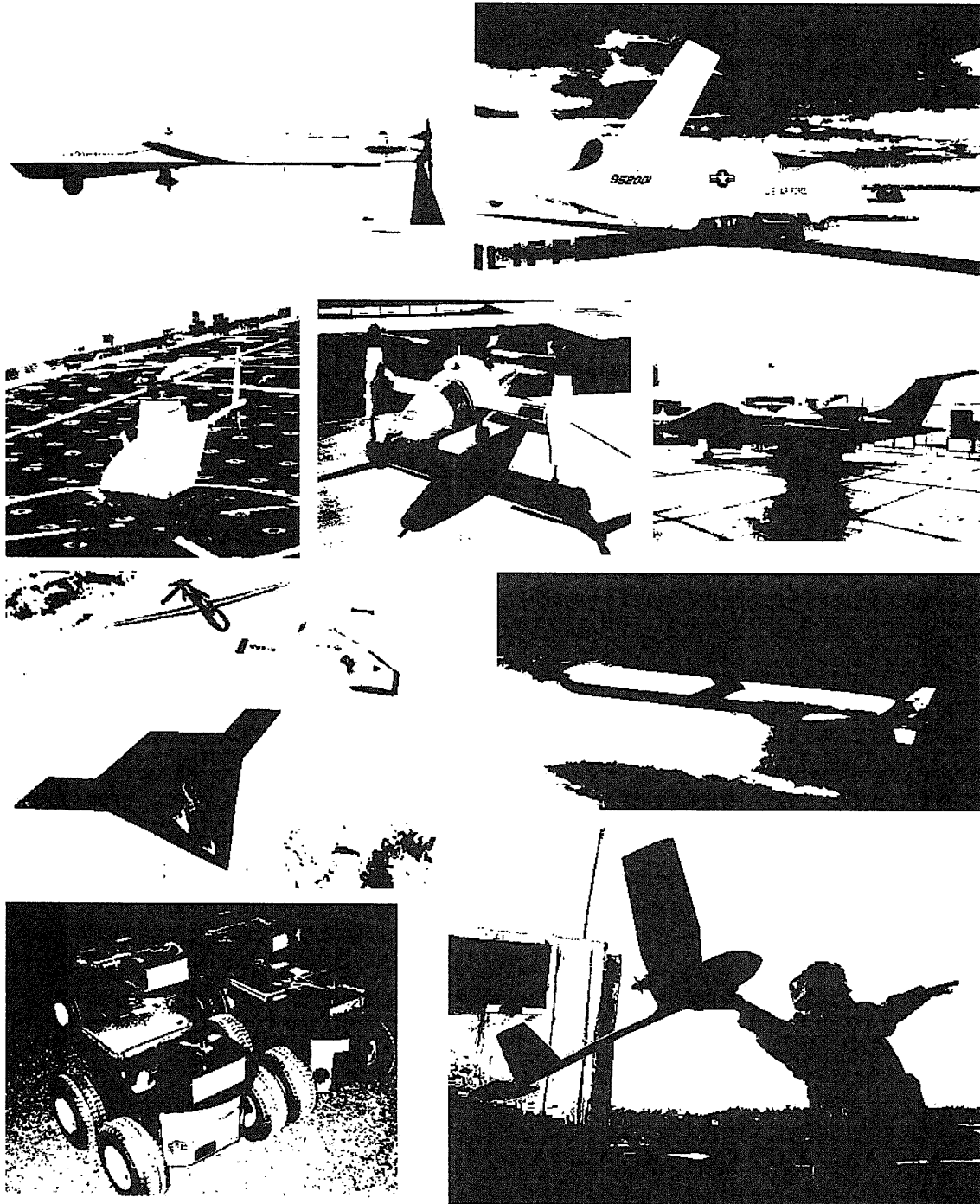
In addition to single vehicle control systems, UAV swarms are being developed. In this scenario a remote pilot executes a command to the swarm which communicates amongst itself to establish, for example, an RF emitter target location.<sup>44</sup> In this type of a control system the raw number of vehicles under one pilot's control can dramatically increase, but the number of swarms then takes the place of the number of vehicles in developing big picture cognitive limits. Such systems are also under development for space exploration.<sup>45</sup>

In 2005, the US Army was operating two tactical surveillance UAVs: the Hunter and its newer replacement, the Shadow. Each UAV requires a team of two operators. Dixon studied the workload of Hunter/Shadow operators and with the help of SMEs designed a simulation to determine if augmentation systems could increase the number of aircraft controlled per pilot from one-half to two. Pilots were responsible for mission completion (reconnaissance of a command target area), locating targets of opportunity (TOO), and on-board system monitoring. There were three levels of pilot aircraft control: baseline, autoalert, and autopilot. In the first two conditions, operators controlled the flight of the aircraft using a joystick to indicate direction; altitude and airspeed were held constant, while a computer controlled the remaining flight parameters (pitch, bank, etc.). Occasionally pilots needed to compensate for wind changes. In the autopilot condition, operators entered the final coordinates of the next command target and the aircraft proceeded in a straight line, compensating automatically for wind changes.

Pilots flew 10 straight flight legs. At the beginning of each leg, the command target was identified and instructions on what to locate were given. If the pilot forgot the instructions, they could hit a "repeat" button. At the end of each leg, high-workload tasks of loitering and zoom/pan the onboard camera to visualize the entire command target were executed. Along each leg between command targets, pilots were instructed to search for TOOs. Primary task completion included locating all relevant information about the command target. Secondary task completion included TOO identification and monitoring for an on-board system failure. The autoalert condition detected system failures and produced an audio alert when the command target was reached.

Results showed that the autoalert augmentation dramatically decreased the time to locate system failures, as well as significantly decreasing the number of requested instruction repeats. Also, the autopilot augmentation dramatically decreased both flightpath deviation and the requested number of repeats, and also dramatically increased the number of TOO detections. Results were similar for the single and dual aircraft scenarios, though some performance, notably TOO detection (92% to 79%), did

decrease between single and dual control cases. The authors attribute this to the UAV interface complexity of 4 screens per vehicle. The authors conclude that further study is required in augmentation strategies and system development.<sup>46</sup>



**Figure 6. Examples of Unmanned Military vehicles:** (left to right, top to bottom) Predator, Global Hawk, Fire Scout, Bell Eagle, BAE Mantis, rendering of an airwing of mixed UAVs, Tomahawk cruise missile, remote ground vehicles, and the Raven.

Lee has previously shown that multiple automated ground vehicles can be autopiloted successfully in a two-stage process when nominal information is available beforehand about the environment to be searched. The two-stage process includes an offline, permission routing table generation where the primary path of each vehicle is planned and downloaded, and an online traffic control stage where small changes to the plan are executed to avoid collisions and deal with contingent activity.<sup>47</sup>

Attempting to increase the limits of UAV to pilot ratio, Ruff examined simulations using three augmented control techniques: manual, management by consent, and management by exception. The authors concluded that the middle level of automation performed best when considering that augmentation algorithms may have associated errors. They also concluded that the absolute maximum number of UAVs a person could control is four.

Cummings addresses the UAV interface issue in a study on retargeting multiple in-flight cruise missiles. Cruise missiles were chosen because they require minimal active piloting. A dual screen interface was constructed very similar to the ATC setup on one map and one list of objects being tracked (in the case of the ATC system, the list is physical rather than a second computer display – see Figure 3b). Cummings side-steps the issue of a cognitive workload metric based on complexity and number of tracked objects by assuming that an operator can execute changes to only one vehicle at a time, and then counting the ratio of time busy making changes to total time in a scenario. This measure is called *utilization* and previous work in systems engineering<sup>48</sup> and queuing theory has shown that utilization rates around 70% max out the typical human operator's ability to hold a big picture. We note that this scenario involves minimal interaction between the missiles, such as flight path conflicts. The missile study is compared to free flight ATC task, where en route and conflict resolution is the responsibility of pilots rather than ATC operators.

Cummings develops several performance measures that are worthy, but for the current treatise an analogue is more succinct: the utilization measure in multiple-object tracking and retargeting is similar to a grandmaster playing many games of chess simultaneously. They walk from board to board, think for a bit, make a move, and then move to the next board. Many high-level chess players can look at board position and evaluate what the next move is without knowledge of previous moves in the game. The question at hand is, how many simultaneous games can the grandmaster play before he is forced to revert to cold position analysis with every new presentation of a game? The conclusion is 16, and it agrees with previous work on free flight ATC.<sup>49</sup> Cummings additionally takes issue with the Ruff limit of four UAVs, noting that this previous experiment included a far more demanding piloting task.

More recent work by Cummings added aircraft heterogeneity to the experiment.<sup>50</sup> She concluded again that 70% utilization is optimal, but notes that the queuing theory concept of *wait times* will significantly affect the maximum number of vehicles that can be attended. In a multiple-vehicle control situation, a vehicle that has exhibited some decrement in performance has an interaction time with the operator to bring it back to acceptable performance. It will then follow its automated routine for a period, called the neglect time, until it falls again below performance threshold and requires the pilot's attention: the time between the need for attention and the beginning of the next interaction time is the wait time. Including wait times generated by more complex



interfaces reduces the maximum number of aircraft controllable to seven.<sup>51</sup> A final study looks at an abstract model of continuous re-planning in different time intervals. This study observes that subjects reacted differently, but in three groups, to automated suggestions for re-planning. The authors conclude that human-automation consensus is the primary driver of system performance.<sup>52</sup> In other words, the human-task interaction must be taken into account, as suggested above by Averty, Loft, and others outlined above.

## Chapter 5: Discussion

We have insight into the maximum number of tracked objects in a multiple space vehicle piloting experiment: it depends greatly on the complexity of the piloting and mission tasks. Brookings showed that in the ATC task, mild complexity affected performance even for a little as six planes being tracked by professional controllers, whereas these controllers regularly track up to ten. Ruff showed that when there is uncertainty in the augmentation system, a maximum of four craft can be controlled and tasked to complete missions. The number of four is consistent with standard estimates of human working memory being able to handle three to five disparate objects at a time. This implies that disparate, complex interfaces require resources from working memory to prevent loss of the big picture.

Augmentation of the human capabilities mainly appears to be helping to maintain a higher number of working memory registers. Whether it is the handwritten blocks for the ATCs, the stored instructions for the Dixon study, or the dual displays of Cummings, the most effective augmentations in the studies above hold information for quick visual retrieval that the brain would otherwise keep in working memory.

Any external automation system to assist the operator in making decisions will have an associated error rate. It was also shown in the ATC and piloting tasks that alerts need to contain a level of noise (false alarms) of 20-25% to avoid automation bias.

Regarding where the future of this work is headed, it is certain that the field is just getting started. Apollo spacecraft required dozens of ground operators to monitor for system failures, and just a few years ago it required two soldiers to operate a simple reconnaissance drone (most of them still do). It is fortunate that ATC and UAV control appear to be extremely applicable to the initial direction of remote space vehicle operations. The 5-year timeframe should see spacecraft-specific simulator studies begin to appear in major peer-reviewed journals.

The major advance to come in developing augmented human capability to pilot multiple spacecraft will be in understanding the cognitive organization of multitasking. With brain imaging it has been shown that multiple resource theory seems to follow the anatomical organization of the brain. In the next 40 years we will find out why the functional studies in multiple task completion don't seem to follow the predictions of multiple resource theory.

## Chapter 6: Conclusions

There is a lack of research in the area of cognitive limits on the number of spacecraft one pilot could control given any mission scenario. Two models for examining what are similar activities are air traffic control and remote piloting of multiple unmanned vehicles. We have shown the research progress in both areas, and the cognitive limits on the number of craft that can be simultaneously controlled are 16 for simple destination selection, 7 for moderately complex piloting and/or mission task completion, and 4 for complex heterogeneous craft. Future research may increase the automation component of aircraft and mission control, but there is no evidence to date that a complete mental picture can be maintained, even with external working memory augmentation, for more than about 16 objects at one time.

We have additionally shown that physiological variables can be objectively employed to indicate overload. Nominal success has been achieved in classifying physiological states near high workload and thus able to predict and thus possibly prevent overload. We expect this classification will be achieved with near perfect accuracy within five years of specific studies being commenced.

## References

- 1 Hart, S. G. & Staveland, L. E. in *Human Mental Workload*, eds Peter A. Hancock & N Meshkati) 139-183 (North Holland, 1988).
- 2 Hart, S. G. *NASA Task Load Index (TLX): 20 Years Later*, <[http://humansystems.arc.nasa.gov/groups/TLX/downloads/HFES\\_2006\\_Paper.pdf](http://humansystems.arc.nasa.gov/groups/TLX/downloads/HFES_2006_Paper.pdf)> (2006).
- 3 Hart, S. G. & Staveland, L. E. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, <<http://humansystems.arc.nasa.gov/groups/TLX/downloads/NASA-TLXChapter.pdf>> (1987).
- 4 Reid, G. B., Shingledecker, C. A. & Eggemeier, F. T. in *Human Factors Society 25th annual meeting*. 522-526 (Human Factors Society).
- 5 Cummings, M. L. & Mitchell, P. J. Operator scheduling strategies in supervisory control of multiple UAVs. *Aerospace Science and Technology* 11, 339-348, doi:10.1016/j.ast.2006.10.007 (2007).
- 6 Graydon, F. X. *et al.* Visual event detection during simulated driving: Identifying the neural correlates with functional neuroimaging. *Transportation Research Part F-Traffic Psychology and Behaviour* 7, 271-286, doi:10.1016/j.trf.2004.09.006 (2004).
- 7 Kandel, E. R., Schwartz, J. H. & Jessell, T. M. *Principles of neural science*. 4th edn, (McGraw-Hill, Health Professions Division, 2000).
- 8 Jennings, J. R., Stringfellow, J. C. & Graham, M. A comparison of the statistical distributions of beat-by-beat heart rate and heart period. *Psychophysiology* 11, 207-210 (1974).
- 9 Porges, S. W. & Byrne, E. A. Research methods for measurement of heart rate and respiration. *Biol Psychol* 34, 93-130 (1992).
- 10 Mulder, L. J. Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biol Psychol* 34, 205-236 (1992).
- 11 Steptoe, A. & Sawada, Y. Assessment of baroreceptor reflex function during mental stress and relaxation. *Psychophysiology* 26, 140-147 (1989).
- 12 Genik, R. J., 2nd, Green, C. C., Graydon, F. X. & Armstrong, R. E. Cognitive avionics and watching spaceflight crews think: generation-after-next research tools in functional neuroimaging. *Aviat Space Environ Med* 76, B208-212 (2005).
- 13 Warm, J. S., Parasuraman, R. & Matthews, G. Vigilance requires hard mental work and is stressful. *Hum Factors* 50, 433-441 (2008).
- 14 Dussault, C., Jouanin, J. C., Philippe, M. & Guezennec, C. Y. EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviat Space Environ Med* 76, 344-351 (2005).
- 15 Kramer, A. F., Trejo, L. J. & Humphrey, D. Assessment of mental workload with task-irrelevant auditory probes. *Biol Psychol* 40, 83-100, doi:0301-0511(95)05108-2 [pii] (1995).
- 16 Backs, R. W. & Walrath, L. C. Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Appl Ergon* 23, 243-254, doi:000368709290152L [pii] (1992).
- 17 Freedman, L. W. *et al.* The relationship of sweat gland count to electrodermal activity. *Psychophysiology* 31, 196-200 (1994).

- 18 Mitchell, J. S., Lowe, T. E. & Ingram, J. R. Rapid ultrasensitive measurement of salivary cortisol using nano-linker chemistry coupled with surface plasmon resonance detection. *Analyst* 134, 380-386, doi:10.1039/b817083p (2009).
- 19 Billings, C. E. & Reynard, W. D. Human factors in aircraft incidents: results of a 7-year study. *Aviat Space Environ Med* 55, 960-965 (1984).
- 20 Wickens, C. D., Mavor, A. S., McGee, J. & National Research Council (U.S.). Panel on Human Factors in Air Traffic Control Automation. *Flight to the future: human factors in air traffic control*. (National Academy Press, 1997).
- 21 Halford, G. S., Wilson, W. H. & Phillips, S. Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behav Brain Sci* 21, 803-831; discussion 831-864 (1998).
- 22 Boag, C., Neal, A., Loft, S. & Halford, G. S. An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics* 49, 1508-1526, doi:10.1080/00140130600779744 (2006).
- 23 Edwards, E. in *Proceedings of British Airline Pilots Association Technical Symposium*. 21-36 (British Airline Pilots Association).
- 24 Chang, Y. H. & Yeh, C. H. Human performance interfaces in air traffic control. *Appl Ergon* 41, 123-129, doi:10.1016/j.apergo.2009.06.002 (2010).
- 25 Di Nocera, F., Fabrizi, R., Terenzi, M. & Ferlazzo, F. Procedural errors in air traffic control: effects of traffic density, expertise, and automation. *Aviat Space Environ Med* 77, 639-643 (2006).
- 26 Lamoureux, T. The influence of aircraft proximity data on the subjective mental workload of controllers in the air traffic control task. *Ergonomics* 42, 1482-1491 (1999).
- 27 Melton, C. E., Smith, R. C., McKenzie, J. M., Wicks, S. M. & Saldivar, J. T. Stress in air traffic personnel: low-density towers and flight service stations. *Aviat Space Environ Med* 49, 724-728 (1978).
- 28 Peiris, M. T. et al. Identification of vigilance lapses using EEG/EOG by expert human raters. *Conf Proc IEEE Eng Med Biol Soc* 6, 5735-5737, doi:10.1109/IEMBS.2005.1615790 (2005).
- 29 Donald, F. M. The classification of vigilance tasks in the real world. *Ergonomics* 51, 1643-1655, doi:10.1080/00140130802327219 (2008).
- 30 Brookings, J. B., Wilson, G. F. & Swain, C. R. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology* 42, 361-377 (1996).
- 31 Collet, C., Averty, P. & Dittmar, A. Autonomic nervous system and subjective ratings of strain in air-traffic control. *Appl Ergon* 40, 23-32, doi:10.1016/j.apergo.2008.01.019 (2009).
- 32 Wilson, G. F. & Fisher, F. The use of cardiac and eye blink measures to determine flight segment in F4 crews. *Aviation Space and Environmental Medicine* 62, 959-962 (1991).
- 33 Wilson, G. F. & Russell, C. A. Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Hum Factors* 45, 381-389 (2003).
- 34 Kaber, D. B., Wright, M. C., Prinzel, L. J., 3rd & Clamann, M. P. Adaptive automation of human-machine system information-processing functions. *Hum Factors* 47, 730-741 (2005).
- 35 Mohler, S. R. The human element in air traffic control: aeromedical aspects, problems, and prescriptions. *Aviat Space Environ Med* 54, 511-516 (1983).

- 36 Loft, S., Sanderson, P., Neal, A. & Mooij, M. Modeling and predicting mental workload in en route air traffic control: critical review and broader implications. *Hum Factors* 49, 376-399 (2007).
- 37 Hendy, K. C., Liao, J. & Milgram, P. Combining time and intensity effects in assessing operator information-processing load. *Hum Factors* 39, 30-47 (1997).
- 38 Boone, J. O. Toward the development of a new aptitude selection test battery for air traffic control specialists. *Aviat Space Environ Med* 51, 694-699 (1980).
- 39 Cohen, D., Wherry, R. J., Jr. & Glenn, F. Analysis of workload predictions generated by multiple resource theory. *Aviat Space Environ Med* 67, 139-145 (1996).
- 40 Hancock, P. A. & Szalma, J. L. *Performance under stress*. (Ashgate Pub., 2008).
- 41 Iani, C. & Wickens, C. D. Factors affecting task management in aviation. *Hum Factors* 49, 16-24 (2007).
- 42 Wickens, C. & Colcombe, A. Dual-task performance consequences of imperfect alerting associated with a cockpit display of traffic information. *Hum Factors* 49, 839-850 (2007).
- 43 Lansdown, T. C., Brook-Carter, N. & Kersloot, T. Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. *Ergonomics* 47, 91-104, doi:10.1080/00140130310001629775 (2004).
- 44 Pack, D. J., Delima, P., Toussaint, G. J. & York, G. Cooperative control of UAVs for localization of intermittently emitting mobile targets. *IEEE Trans Syst Man Cybern B Cybern* 39, 959-970, doi:10.1109/TSMCB.2008.2010865 (2009).
- 45 Wang, J., Qu, Z. H., Ihlefeld, C. M. & Hull, R. A. A control-design-based solution to robotic ecology: Autonomy of achieving cooperative behavior from a high-level astronaut command. *Autonomous Robots* 20, 97-112, doi:10.1007/s10514-006-5942-5 (2006).
- 46 Dixon, S. R., Wickens, C. D. & Chang, D. Mission control of multiple unmanned aerial vehicles: a workload analysis. *Hum Factors* 47, 479-487 (2005).
- 47 Lee, J. H., Lee, B. H. & Choi, M. H. A real-time traffic control scheme of multiple AGV systems for collision free minimum time motion: A routing table approach. *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans* 28, 347-358 (1998).
- 48 Rouse, W. B. *Systems engineering models of human-machine interaction*. (North Holland, 1980).
- 49 Cummings, M. L. & Guerlain, S. Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors* 49, 1-15 (2007).
- 50 Kornguth, S. E., Steinberg, R. & Matthews, M. D. *Neurocognitive and physiological factors during high-tempo operations*. (Ashgate, 2010).
- 51 Cummings, M. L. & Mitchell, P. J. Predicting controller capacity in supervisory control of multiple UAVs. *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans* 38, 451-460, doi:10.1109/tsmca.2007.914757 (2008).
- 52 Cummings, M. L., Clare, A. & Hart, C. The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling. *Human Factors* 52, 17-27, doi:10.1177/0018720810368674 (2010).