Triggering Adaptive Automation in Naval Command and Control

Tjerk de Greef ¹,² and Henryk Arciszewski ²

¹Delft University of Technology, ²TNO Defence, Security and Safety
The Netherlands

1. Introduction

In many control domains (plant control, air traffic control, military command and control) humans are assisted by computer systems during their assessment of the situation and their subsequent decision making. As computer power increases and novel algorithms are being developed, machines move slowly towards capabilities similar to humans, leading in turn to an increased level of control being delegated to them. This technological push has led to innovative but at the same time complex systems enabling humans to work more efficiently and/or effectively. However, in these complex and information-rich environments, task demands can still exceed the cognitive resources of humans, leading to a state of overload due to fluctuations in tasks and the environment. Such a state is characterized by excessive demands on human cognitive capabilities resulting in lowered efficiency, effectiveness, and/or satisfaction. More specifically, we focus on the human-machine adaptive process that attempts to cope with varying task and environmental demands.

In the research field of adaptive control an adaptive controller is a controller with adjustable parameters and a mechanism for adjusting the parameters (Astrom & Wittenmark, 1994, p. 1) as the parameters of the system being controlled are slowly time-varying or uncertain. The classic example concerns an airplane where the mass decreases slowly during flight as fuel is being consumed. More specifically, the controller being adjusted is the process that regulates the fuel intake resulting in thrust as output. The parameters of this process are adjusted as the airplane mass decreases resulting in less fuel being injected to yield the same speed.

In a similar fashion a human-machine ensemble can be considered an adaptive controller. In this case, human cognition is a slowly time-varying parameter, the adjustable parameters are the task sets that can be varied between human and machine, and the control mechanism is an algorithm that ‘has insight’ in the workload of the human operator (i.e., an algorithm that monitors human workload). Human performance is reasonably optimal when the human has a workload that falls within certain margins; severe performance reductions result from a workload that is either too high or (maybe surprisingly) too low. Consider a situation where the human-machine ensemble works in cooperation in order to control a process or situation. Both the human and the machine cycle through an information processing loop, collecting data, interpreting the situation, deciding on actions to achieve one or more stated goals and acting on the decisions (see for example Coram, 2002;
Parasuraman et al., 2000). If the human is getting overloaded, the control mechanisms should adjust the parameters that regulate the balance of work between human and machine and work should be reallocated to the machine in order to lower the cognitive burden of the human and optimize the performance of the human machine ensemble. Of course we must be able to automate some or all of the loop so that work can indeed be delegated to the machine. And humans must be willing to delegate the responsibility as well. The process of reallocation of the workload between man and machine is referred to as adaptive automation.

Adaptive automation is based on the idea of supporting the human only at those moments when its performance is in jeopardy. W. B. Rouse (1988) introduced adaptive aiding as an initial type of adaptive automation. Rouse stated that adaptive aiding is a human-machine system-design concept that involves using aiding/automation only at those points in time when human performance needs support to meet operational requirements (Rouse, 1988, p. 431). Whether one uses the terms adaptive automation, dynamic task allocation, dynamic function allocation, or adaptive aiding, they all reflect the dynamic reallocation of work in order to improve human performance or to prevent performance degradation. As a matter of fact, adaptive automation should scale itself down when things become quieter again and the goal of adaptive automation could be stated as trying to keep the human occupied within a band of ‘proper’ workload (see Endsley & Kiris, 1995). Periods of ‘underload’ can have equally disastrous consequences as periods of overload due to slipping of attention and loss of situational awareness. A number of studies have shown that the application of adaptive automation enhances performance, reduces workload, improves situational awareness, and maintains skills that are deteriorating as a consequence of too highly automated systems (Bailey et al., 2006; Hilburn et al., 1997; Inagaki, 2000a; Kaber & Endsley, 2004; Moray et al., 2000; Parasuraman et al., 1996; Scallen et al., 1995).

One of the challenging factors in the development of successful adaptive automation concerns the question of when changes in the level of automation must be effectuated. The literature repository utilizes the idea of ‘the workload being too high or too low’ as a reason to trigger the reallocation of work between the human and the machine. At the same time it acknowledges the fact that it remains difficult to give the concept a concrete form. We simply state that workload measurements of some sort are required in order to optimize the human-machine performance. Performance measurements are one way to operationalize such workload measurements and the next section discusses the various strategies in detail.

2. Previous Work

The success of the application of adaptive automation depends in part on the quality of the automation and the support it offers to the human. The other part constitute when changes in the level of automation are effectuated. ‘Workload’ generally is the key concept to invoke such a change of authority. Most researchers, however, have come to the conclusion that workload is a multidimensional, multifaceted concept that is difficult to define. It is generally agreed that attempts to measure workload relying on a single representative measure are unlikely to be of use (Gopher & Donchin, 1986). The definition of workload as an intervening variable similar to attention that modulates or indexes the tuning between the demands of the environment and the capacity of the operator (Kantowitz, 1987) seems to capture the two main aspects of workload, i.e., the capacity of humans and the task demands made on them. The workload increases when the capacity decreases or the task demands increase. Both capacity and task demands
are not fixed entities and both are affected by many factors. Skill and training, for example, are two factors that increase capacity in the long run whereas capacity decreases when humans become fatigued or have to work under extreme working conditions for a prolonged period.

If measuring workload directly is not a feasible way to trigger the adaptive automation mechanism, other ways must be found. Wilson and Russell (2007) define five strategies based on a division by Parasuraman et al (1996). They state that triggers can be based on critical events, operator performance, operator physiology, models of operator cognition, and hybrid models that combine the other four techniques. The workload perceived by the human himself or by a colleague may lead to an adaptation as well, although in such a case some papers refrain from the term adaptive automation and utilize ‘adaptable automation,’ as the authority shift is not instigated by the automated component. Against the first option (operator indicates a workload that is too high or too low that in turn results in work adjustments) counts the fact that he or she is already over or underloaded and the additional switching task would very likely be neglected. The second option therefore seems more feasible, but likely involves independent measurements of workload to support the supervisor’s view, leading to a combination of the supervision method and other methods.

The occurrence of critical events can be used to change to a new level of automation. Critical events are defined as incidents that could endanger the goals of the mission. Scerbo (1996) describes a model where the system continuously monitors the situation for the appearance of critical events and the occurrence of such an event triggers the reallocation of tasks. Inagaki has published a number of theoretical models (Inagaki, 2000a; Inagaki, 2000b) where a probabilistic model was used to decide who should have authority in the case of a critical event.

A decline in operator performance is widely regarded as a potential trigger. Such an approach measures the performance of the human over time and regards the degradation of the performance as an indication of a high workload. Many experimental studies derive operator performance from performance measurements of a secondary task (Clamann et al., 2002; Kaber et al., 2006; Kaber & Riley, 1999; Kaber et al., 2005). Although this approach works well in laboratory settings, the addition of an artificial task to measure performance in a real-world setting is unfeasible so extracting performance measures from the execution of the primary task seems the only way to go.

Physiological data from the human are employed in various studies (Bailey et al., 2006; Byrne & Parasuraman, 1996; Prinzel et al., 2000; Veltman & Gaillard, 1998; Wilson & Russell, 2007). The capability of human beings to adapt to variable conditions, however, may distort accurate measurements (Veltman & Jansen, 2004). There are two reasons why physiological measures are difficult to use in isolation. First of all, the human body responds to an increased workload in a reactive way. Physiological measurements therefore provide the system with a delayed cognitive workload state of the operator instead of the desired real-time measure. Second, it is possible that physiological data indicate high workload but that these not necessarily commensurate with poor performance. This is the case when operators put in extra effort to compensate for increases in task demands. At least several measurements (physiological or otherwise) are required to get rid of such ambiguities.

The fourth approach uses models of operator cognition. These models are approximations of human cognitive processes for the purpose of prediction or comprehension of human operator state and workload. The winCrew tool (Archer & Lockett, 1997), for example,
implements the multiple resource theory (Wickens, 1984) to evaluate function allocation strategies by quantifying the moment-to-moment workload values. Alternatively, the human’s interactions with the machine can be monitored and evaluated against a model to determine when to change levels of automation. In a similar approach, Geddes (1985) and Rouse, Geddes, and Curry (1987) base adaptive automation on the human’s intentions as predicted from patterns of activity.

The fifth approach follows Gopher and Donchin (1986) in that a single method to measure workload is too limited. Hybrid models therefore combine a number of triggering techniques because the combination is more robust against the ambiguities of each single model.

Each of the five described approaches has been applied more or less successfully in an experimental setting, especially models that consider the effects of (neuro)physiological triggers and critical events. Limited research is dedicated to applying a hybrid model that integrates operator performance models and models of operator cognition. We have based our trigger model on precisely such a combination because we feel our approach to adaptive automation using an object-oriented model (de Greef & Arciszewski, 2007) offers good opportunities for an operational implementation. The cognitive model we use is based in turn on the cognitive task load (CTL) model of Neerincx (2003). In addition, we provide a separate mechanism for critical events.

3. Naval Command and Control

As our implementation domain concerns naval command and control (C2), we begin our discussion with a brief introduction to this subject. Specifically, command and control is characterized as focusing the efforts of a number of entities (individuals and organizations) and resources, including information, toward the achievement of some task, objective, or goal (Alberts & Hayes, 2006, p. 50). These activities are characterized by efforts to understand the situation and subsequently acting upon this understanding to redirect it toward the intended one. A combat management system (CMS) supports the team in the command center of a naval vessel with these tasks. Among other things this amounts to the continuous execution of the stages of information processing (data collection, interpretation, decision making, and action) in the naval tactical domain and involves a number of tasks like correlation, classification, identification, threat assessment, and engagement. Correlation is the process whereby different sensor readings are integrated over time to generate a track. The term track denotes the representation of an external platform within the CMS, including its attributes and properties, rather than its mere trajectory. Classification is the process of determining the type of platform of a track and the identification process attempts to determine its identity in terms of it being friendly, neutral, or hostile. The threat assessment task recognizes entities that pose a threat toward the commanded situation. In other words, the threat assessment task assesses the danger a track represents to the own ship or other friendly ships or platforms. One should realize that hostile tracks do not necessarily imply a direct threat. The engagement task includes the decision to apply various levels of force to neutralize a threat and the execution of this decision. Because the identification process uses information about such things as height, speed, maneuvering, adherence to an air or sealane, and military formations, there is a continuous need to monitor all tracks with respect to such aspects. Therefore monitoring is also part of the duties of a command team. See Figure 1 for an overview of C2 tasks in relation to a track.
4. The Object-oriented framework

Before describing triggering in an object-oriented framework, we summarize our previous work (Arciszewski et al., in press).

4.1 Object-Oriented Work Allocation

We have found it fruitful to focus on objects rather than tasks in order to distribute work among actors (compare Bolderheij, 2007, pp. 47-48). Once we have focused our attention on objects, tasks return as the processes related to the objects (compare Figure 1). For example, some of the tasks that can be associated with the all-evasive ‘track’ object in the C2 domain are classification, the assignment of an identity, and continuous behavioral monitoring (compare Figure 1). The major advantage of the object focus in task decomposition is that it is both very easy to formalize and comprehensible by the domain users. Partitioning work using tasks only has proven difficult. If we consider identification, for example, this task is performed for each object (track) in turn.

Figure 1. Some of the more important tasks a command crew executes in relation to a track

4.2 Concurrent Execution and Separate Work Spaces

Instead of letting a task be performed either by the human or the machine, we let both parties do their job concurrently. In this way both human and machine arrive at their own interpretation of the situation, building their respective world-views (compare Figure 2). One important result of this arrangement is the fact that the machine always calculates its view, independent of whether the human is dealing with the same problem or not. To allow this, we have to make provisions for ‘storage space’ where the two parties can deposit the information pertaining to their individual view of the world. Thus we arrive at two separate data spaces where the results of their computational and cognitive efforts can be stored. This has several advantages. Because the machine view is always present, advice can be readily looked up. Furthermore, discrepancies between the two world views can lead to warnings from the machine to the human that the latter’s situational awareness may no longer be up to date and that a reevaluation is advisable. Assigning more responsibility to the machine, in practice comes down to the use of machine data in situation assessment, decision making, and acting without further intervention from the human.
Figure 2. The two different world views and a comparison of them by the system. A difference between the interpretation of the two worlds could lead to an alert of the human.

4.3 Levels of Automation

Proceeding from the machine and human view, levels of automation (LoA) more or less follow automatically. Because the machine view is always available, advice is only a key press or mouse click away. This readily available opinion represents our lowest LoA (ADVICE). At the next higher LoA, the machine compares both views and signals any discrepancies to the human, thus alerting the user to possible gaps or errors in his situational picture. This signalling functionality represents our second LoA (CONSENT). At the higher levels of automation we grant the machine more authority. At our highest LoA (SYSTEM), the machine entirely takes over the responsibility of the human for certain tasks. At the lower LoA (VETO), the machine has the same responsibility, but alerts the human to its actions, thus allowing the latter to intervene.

Adaptive automation now becomes adjusting the balance of tracks for each task between the human and the machine. By decreasing the number of tracks under control of the human, the workload of the human can be reduced. Increasing the number of tracks managed by the human on the other hand results in a higher workload.

5. Global and local adaptation

Having outlined an architectural framework for our work, we now focus on the problem of triggering. We envision two clearly different types of adaptation. The distinction between the two types can be interpreted as that between local and global aiding (de Greef & Lafeber, 2007, pp. 68-69). Global aiding is aimed at the relief of the human from a temporary overload situation by taking over parts of the work. If on the other hand the human misses a specific case that requires immediate attention in order to maintain safety, local aiding comes to the rescue. In both cases work is shifted from the human to the machine, but during global aiding this is done in order to avoid the overwhelming of the human, whereas local aiding offers support in those cases the human misses things. As indicated before, global aiding should step back when things become quiet again in order to keep the human within a band of ‘proper’ workload (see Endsley & Kiris, 1995). On the other hand, a human is not overloaded in cases where local adaptation is necessary; he or she may be just missing those
particular instances or be postponing a decision with potentially far-reaching consequences. A further distinction is that local aiding concerns itself with a specific task or object whereas global aiding takes away from the operator that work that is least detrimental to his or her situational awareness. According to this line of reasoning a local case ought to be an exception and the resulting actions can be regarded as a safety net. The safety net can be realized in the form of separate processes that check safety criteria. In an ideal world, global adaptation would ensure that local adaptation is never necessary because the human always has enough cognitive resources to handle problems. But things are not always detected in time and humans are sometimes distracted or locked up so that safety nets remain necessary.

6. Triggering local aiding

Local aiding is characterized by a minimum time left for an action required to maintain safety and be able to achieve the mission goals. Activation of such processing is through triggers that are similar to the critical events defined by Scerbo (1996). The triggers are indicators of the fact that a certain predefined event that endangers mission goals is imminent and that action is required shortly. In the case of naval C2 a critical event is usually due to a predefined moment in the (timeline of the) state of an external entity and hence it is predictable to some extent. Typically, local aiding occurs in situations where either the human misses something due to a distraction by another non-related event or entity, to tunnel vision, or to the fact that the entity has so far been unobserved or been judged to be inconsequential. In the naval command and control domain, time left as a way to initiate a local aiding trigger can usually be translated to range from the ship or unit to be protected. In most cases therefore triggers can be derived from the crossing of some critical boundary. Examples are (hostile) missiles that have not been engaged by the crew at a certain distance or tracks that are not yet identified at a critical range called the identification safety range (ISR). The ship’s weapon envelopes define a number of critical ranges as well. It is especially the minimum range, within which the weapon is no longer usable, that can be earmarked as a critical one.

7. Triggering global aiding

One of the advantages of the object-oriented framework outlined in section 4 is that it offers a number of hooks for the global adaptation approach. The first hook is the difference between human world-view and machine world-view (see sect. 4.2). The second hook is based on the number and the character of the objects present and is utilized for estimating the workload imposed on the human by the environment. In the case of military C2 the total number of tracks provides an indication of the volume of information processing whereas the character of the tracks provides an indication of the complexity of the situation. These environmental items therefore form the basis for our cognitive model.

7.1 The Operator Performance Model

Performance is usually defined in terms of the success of some action, task, or operation. Although many experimental studies define performance in terms of the ultimate goal, real world settings are more ambiguous and lack an objective view of the situation (the ‘ground truth’) that could define whether an action, task, or operator is successful or not. Defining
performance in terms of reaction times is another popular means although some studies found limited value in utilizing performance measures as a single way to trigger adaptive automation. This has been our experience as well (de Greef & Arciszewski, 2007).

As explained in section 4.2, the object-oriented framework includes the principle of separate workspaces for man and machine. This entails that both the machine and the human construct their view of the world and store it in the system. For every object (i.e., track) a comparison between the two world views can then be made and significant differences can be brought to the attention of the human. This usually means that new information has become available that requires a reassessment of the situation as there is a significant chance that the human’s world view has grown stale and his or her expectations may no longer be valid. We use these significant differences in two ways to model performance.

First, an increase in the number of differences between the human world view and the machine world view is viewed as a performance decrease. Although differences will inevitably occur, as the human and the machine do not necessarily agree, an increasing skew between the two views is an indication that the human has problems with his or her workload. Previous work suggested that the subjective workload fluctuated in proportion to the density of signals resulting from skew differences (van Delft & Arciszewski, 2004). The average reaction time to these signals is used as a second measure of performance. Utilizing either skew or reaction times as the only trigger mechanism is problematic because of the sparseness of data due to the small number of significant events per time unit in combination with a wide spread of reaction times (de Greef & Arciszewski, 2007). The combined use of skew and reaction times provides more evidence in terms of human cognitive workload. This in turn is enhanced by the operator cognitive model discussed below.

7.2 The Operator Cognition Model

While the operator performance model is aimed to get a better understanding of the human response to the situation, the operator cognition model aims at estimating the cognitive task load the environment exerts on the human operator. The expected cognitive task load is based on Neerincx’s (2003) cognitive task load (CTL) model and is comprised of three factors that have a substantial effect on the cognitive task load.

The first factor, percentage time occupied (TO), has been used to assess workload for time-line assessments. Such assessments are based on the notion that people should not be occupied more than 70 to 80 percent of the total time available. The second load factor is the level of information processing (LIP). To address cognitive task demands, the cognitive load model incorporates the skill-rule-knowledge framework of Rasmussen (1986) where the knowledge-based component involves the highest workload. To address the demands of attention shifts, the model distinguishes task-set switching (TSS) as a third load factor. It represents the fact that a human operator requires time and effort to reorient himself to a different context. These factors present a three-dimensional space in which all human activities can be projected as a combined factor (i.e., it displays the workload due to all activities combined). Specific regions indicate the cognitive demands activities impose on a human operator. Figure 3 displays the three CTL factors and a number of cognitive states.

Applying Neerincx’s CTL model leads to the notion that the cognitive task load is based on the volume of information processing (reflecting time occupied), the number of different objects and tasks (task set switching), and the complexity of the situation (level of information
Figure 3. The three dimensions of Neerincx’s (2003) cognitive task load model: time occupied, task-set switches, and level of information processing. Within the cognitive task load cube several regions can be distinguished: an area with an optimal workload displayed in the center, an overload area displayed in top vertex, and an underload area displayed in the lower vertex.

The second CTL factor is the task set switching factor. We recognize two different types of task set switching, each having a different effect size $C_x$. The human operator can change between tasks or objects (tracks). The first switch relates to the attention shift that occurs as a consequence of switching tasks, for example from the classification task to the engagement task. The second type of TSS deals with the required attention shift as a result of switching from object to object. The latter type of task switch is probably cognitively less demanding because it is associated with changing between objects in the same task and every object has similar attributes, each requiring similar information-processing capabilities.

Finally, a command and control context can be expressed in terms of complexity (i.e., LIP). The LIP of an information element in C2, a track, depends mainly on the identity of the track. For example, ‘unknown’ tracks result in an increase in complexity since the human operator has to put cognitive effort in the process of ascertaining the identity of tracks of which relatively little is known. The cognitive burden will be less for tracks that are friendly or neutral.

The unknown, suspect, and hostile tracks require the most cognitive effort for various reasons. The unknown tracks require a lot of attention because little is known about them and the operator will have to ponder them more often. On the other hand, hostile tracks require considerable cognitive effort because their intent and inherent danger must be decided. Especially in current-day operations, tracks that are labeled hostile do not necessarily attack and neutralization might only be required in rare cases of clear hostile intent. Suspect tracks are somewhere between hostile and unknown identities, involving too little information to definitely identify them and requiring continuous threat assessment as well. We therefore conclude a relationship between the LIP, an effect size $C$, and the numbers of hostile,
suspect, and unknown tracks and the other categories where the effect is larger for the hostile, suspect, and unknown tracks.

7.3 The hybrid cognitive task load model
The operator performance model describes a relation between performance and 1) average response time and 2) skew between the human view and the machine view of the situation. A decrease in performance, in its turn, is the result of a task load that is too high (see de Greef & Arciszewski, 2007). In the second place, the model of operator cognition describes a relation between the environment and the cognitive task load in terms of the three CTL factors. We therefore define a relation between cognitive task load and the number of tracks \((N_T)\) the number of objects \((N_O)\), and the number of difficult tracks \((N_{U\cup S\cup H})\).

In all cases, a further investigation into the relation between the cognitive task load indicators and the performance measurements is worthwhile. We expect that a change in one of the workload indicators \(N_T, N_O, N_{U\cup S\cup H}\) results in a change in cognitive load, leading in turn to a (possibly delayed) change in performance and hence a change in a performance measurement.

8. Experiment I
In order to see whether the proposed model of operator cognition is a true descriptor for cognitive workload we looked at data from an experiment. This experiment investigated the relation between the object-oriented approach and cognitive task load. More specifically, this experiment attempted to answer the question whether CTL factors properly predict or describe changes in cognitive workload.

8.1 Apparatus & Procedure
The subjects were given the role of human operators of (an abstracted version of) a combat management workstation aboard naval vessels. The workstation comprised a schematic visual overview of the nearby area of the ship on a computer display, constructed from the data of radar systems. On the workstation the subject could manage all the actions required to achieve mission goals. Before the experiment, the subjects were given a clear description of the various tasks to be executed during the scenarios. Before every scenario, a description about the position of the naval ship and its mission was provided. The experiment was conducted in a closed room where the subjects were not disturbed during the task. During the experiment, an experimental leader was situated roughly two meters behind the subject to assist when necessary.

8.2 Participants
Eighteen subjects participated in the experiment and were paid EUR 40 to participate. The test subjects were all university students, with a good knowledge of English. The participant group consisted of ten men and eight women. They had an average age of 25, with a standard deviation of 5.1.
8.3 Experimental tasks
The goal of the human operator during the scenarios was to monitor, classify, and identify every track (i.e. airplanes and vessels) within a 38 nautical miles range around the ship. Furthermore, in case one of these tracks showed hostile intent (in this simplified case a dive toward the ship), they were mandated to protect the naval vessel and eliminate the track. To achieve these goals, the subject was required to perform three tasks. First, the classification task gained knowledge of the type of the track and its properties using information from radar and communication with the track, air controller, and/or the coastguard. The subject could communicate with these entities using chat functionality in the CMS. The experimental leader responded to such communications. The second task was the identification process that labeled a track as friendly, neutral, or hostile. The last task was weapon engagement in case of hostile intent as derived from certain behavior. The subject was required to follow a specific procedure to use the weapons.

8.4 Scenarios
There were three different scenarios, each implying a different cognitive task load. The task loads were under-load, normal load, and an overload achieved by manipulating two of the three CTL factors. First, the total number of tracks in a scenario was changed. If many tracks are in the observation range, the percentage of the total time that the human is occupied is high (see section 7.2). Second, a larger amount of tracks that show special behavior and more ambiguous properties increases the operator’s workload. It forces the human operator to focus attention and to communicate more in order to complete the tasks.

We hypothesize that manipulation of these two items has an effect on the cognitive task load factors, similar to our model of operator cognition described in section 7.2. In summary:
- Time occupied: manipulated by the number of tracks in the range of the ship.
- Task set switches: likewise manipulated by number of tracks in the range.
- Level of information processing: manipulated by the behavior of the tracks.

Table 1 provides the values used per scenario. The scenarios were presented to the participants using a Latin square design to compensate for possible learning effects. The TO, TSS, and LIP changes were applied at the same time.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Total number of tracks within 38 nautical miles</th>
<th>Track with hostile behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under-load scenario</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Normal workload scenario</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Overload scenario</td>
<td>34</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 1. Total number of tracks and the number of tracks with hostile behavior per scenario

8.5 Results
In order to verify whether the manipulated items affected the load factors and induced mental workload as expected, the subjects were asked to indicate their workload. Every 100 seconds subjects had to rate his or her perceived workload on a Likert scale (one to five). Level 1 indicated low workload, level 3 normal workload, and level 5 high workload. The levels in between indicate intermediate levels of workload.
Figure 4. The subjective workload per scenario as indicated every 100 seconds on a five point Likert scale. Note: for the mental workload verification, N = 17 as the data of one subject was missing due to a failure in logging.

Repeated-measures ANOVA reveals a significant effect in perceived cognitive task load between the three scenario’s (F(2,32) = 190.632, p < 0.001, see Figure 4). Least square difference post-hoc analysis reveals that all three means were significantly different (p < 0.05). Compared to the under-load scenario, the perceived mental workload was significantly higher in the normal workload scenario. In turn, the perceived mental workload in the overload scenario was significantly higher again than in the normal-workload scenario.

8.6 Conclusion
The data from the experiment reveal that manipulation of the CTL factors using numbers and types of domain objects has a significant effect on the subjective cognitive task load. We therefore conclude that the total number of tracks and the number of tracks with extraordinary behavior are good indicators of the difficulty the environment poses on a human operator. The data supports our model of operator cognition described in section 7.2.

9. Experiment II
While experiment I studied the relation between the object-oriented approach and cognitive task load in a naive setting, the second experiment investigated the performance model and the application of a hybrid cognitive task load model in a semi-realistic setting of naval operations during peace keeping and embargo missions. Experiment II was in the first place designed to compare the efficiency and effectiveness between an adaptive and a non-adaptive mode of the CMS during high-workload situations. The results revealed a clear performance increase in the adaptive mode with no differentiation in subjective workload and trust (for a detailed review see de Greef et al., 2007). The triggers for the adaptive mode, mandated by the high-workload situations, were mainly based on performance measures and to a lesser extent on cognitive indicators.
In spite of the different goal, the data of the non-adaptive subset of runs help investigating the claims with respect to the proposed hybrid model. In addition to the model of operator cognition, we hypothesize that the operator performance model is a predictor for workload in accordance with section 7.1 and 7.3. Experiment II therefore uses the non-adaptive subset of the data to investigate this aspect.

9.2 Subjects, Tasks and Apparatus
The subjects were four warfare officers and four warfare officer assistants of the Royal Netherlands Navy with several years of operational experience. All subjects were confronted with a workstation called the Basic-T (van Delft & Schraagen, 2004) attached to a simulated combat management system. The Basic-T (see Figure 5) consists of four 19-inch touch screens arranged in a T-shaped layout driven by two heavy-duty PCs. The Basic-T functioned as an operational workstation in the command centre of a naval ship and was connected by means of a high-speed data bus to the simulated CMS running on an equally simulated naval vessel.

In all cases the primary mission goal for the subjects was to build a complete maritime picture of the surroundings of the ship and to defend the ship against potential threats. Building the maritime picture amounted to monitoring the operational space around the vessel and classifying and identifying contacts. The defense of the ship could entail neutralizing hostile entities. As the sensor reach of a modern naval ship extends to many miles around the ship, the mission represented a full-time job. In addition, the subjects were responsible for the short-term navigation of the ship, steering it toward whatever course was appropriate under the circumstances and had a helicopter at their disposal to investigate the surrounding surface area. Although the use of a helicopter greatly extended surveillance capabilities, it also made the task heavier because of the increased data volume and the direction and control of the helicopter.
Each subject was offered an abstract top-down tactical view of the situation in order to support his or her situational awareness. The tactical display was amended by a second display that contained detailed information about the selected track (for example, flight profile, classification, and radar emissions). A chat console aided the subject to gather and distribute information. The subject could communicate with and assign a new task to the helicopter and contact external entities such as the coastguard and aircraft. One of the experimental leaders controlled the helicopter, generally executed commands to emulate on-board personnel (controlling the fire control radar, gunnery, etc.) and responded to the chat.

9.3 Procedure
The subjects participated in the experiment for two days. The first day was divided into two parts. In the first part of the day the participants were informed about the general goals of the experiment and the theoretical background of the research. The second part was used to familiarize the participants with the Basic-T and the various tasks. This stage consisted of an overall demonstration of the system and three training scenarios. The offered scenarios showed an increasing complexity and the last scenario approached the complexity of the experimental trials.

The evaluation took place on the second day. Prior to the experimental trials both subjects were offered a scenario to refresh their memory on the ins and outs of the workstation. After this warming up, the trials commenced. After each run a debriefing session with the subject was held in order to discuss his or her experiences.

9.4 Scenarios
A total number of four scenarios were developed in cooperation with experts of the Royal Netherlands Navy. All scenarios were intended to pose a substantial workload to the subjects and included various threats or suspicious-looking tracks that contributed to the workload. Two of the four scenarios were developed around more or less traditional air and surface warfare in a high-tension situation while the other two scenarios were situated against a civilian background where countering smuggling was the main mission objective. The latter two scenarios were made more ambiguous and threatening by the possibility of a terrorist attack. All scenarios took about 20 minutes to conclude. Because of the relative freedom of the subjects to operate their resources, differences in the actual runs of the scenario occurred. For example, by sending the helicopter to different locations, the actual time at which hostile ships were positively identified could shift by one to two minutes. Generally, however, the scenarios ran in agreement.

In order to exclude sequence effects and minimize effects of learning, increasing acquaintance with the workstation, personal experience, etc., the scenarios were allocated in a balanced way where each subject executed one of each scenario-type.

9.5 Experimental setup
As only the data of the non-adaptive mode were used for this investigation, three independent variables remain: scenario type, subject rank, and scenario time. Scenario type was balanced within subjects, subject rank between subjects, and the scenarios were divided into 16 equal time slots. The start of the first time slot was dependent on the first time a subject entered his or her subjective workload (thereafter every 80 seconds). The rank
variable described whether the subject worked as a warfare officer assistant (Chief Petty officer) or a warfare officer (Lieutenant Commander).

A number of dependent variables was measured:

- The **subjective workload** as rated every 80 seconds during each scenario on a one dimensional Likert rating scale ranging from one to five, one meaning heavy underload and boredom, three a comfortable and sustainable workload and five an overload of the operator. Six was logged in case the subject didn’t indicate his or her subjective workload and was converted to five during the analysis.

- The **number of tracks** and the **number of signals** were logged every second.

- The **performance** in terms of tracks handled and reaction time to signals was logged every second.

- The data describing the **human world-view** and the **machine world-view** was stored (logged every second). This includes the position, class, and identity of each track.

### 9.6 Hypotheses

The data from the experiment enabled us to investigate the claims with respect to the operator performance model and the hybrid model. Software, known as the cognitive task load and performance monitor, was developed both to generate the adaptation triggers during the original experiment (on-line) and to facilitate an off-line first-order analysis between performance and cognitive effects on workload. The CTL monitor visualized the reaction times, the number of tracks, the number of signals, the machine world view, the human world view, and the subjective workload (see Figure 6).

The world views were ‘summarized’ in numbers of friendly, assumed friendly, neutral, suspect, and hostile tracks. For the tracks designated ‘assumed friendly’ and ‘suspect’, not enough hard data are available to assign a definite identity to them, although they ‘seem to be’ friendly and hostile, respectively. Tracks can also be designated ‘unknown’, in which case so little is known about them that they can be anything. As tracks are first observed they are assigned the identity ‘pending’, meaning the operator has not had time to take a look at them yet. A lot of pending tracks is an indication that the user is behind with his or her work (a lack of time). A situation with a lot of ‘unknown’ tracks rather indicates a lack of data instead.

A first order analysis of the data from the experiments using the CTL and performance monitor resulted in the generation of three hypotheses.

1. **Because all scenarios were intended to stress the subjects, the difference between the scenarios was not expected to be large. Nevertheless the smuggling scenarios seemed to contain more ‘theoretically difficult’ tracks (as a percentage of the total number of tracks to compensate for differences in the total number of tracks) compared to the traditional warfare scenarios. The ‘theoretically difficult’ tracks consist of ambiguous, suspect or unknown, tracks as discussed in sect. 7.**

2. **If ‘theoretically difficult’ tracks are experienced by subjects as difficult as well, the smuggling scenarios should show an increased workload when compared to the traditional scenarios.**

3. **The warfare officers seemed to show a different behavior in dealing with the situation compared to the warfare assistants.**
9.7 Statistical Results

For each dependent variable a repeated-measures analysis MANOVA was used to analyze the data using scenario and time as a within factor and subject rank as a between factor. In all cases, an alpha level of .05 was used to determine statistical significance. Post-hoc analyses were conducted using Tukey’s HSD and Fishers LSD tests.
Analysis of the two different scenario types (smuggling vs. traditional) reveals that the smuggling scenarios contain an effect in terms of more ambiguous tracks as compared to the traditional ones \( F(2, 153) = 59.463, p < .0001 \) according to both human and machine interpretation (see Figure 7). The value is expressed as a percentage of the total number of tracks per time unit to compensate for differences in the total number of tracks. Tukey’s post-hoc analysis reveals that the smuggling scenarios have more ‘difficult’ tracks according to the human interpretation of the world \( (p < .01) \) and the machine interpretation of the world \( (p < .0001) \). Detailed analysis of the class of ambiguous tracks discloses that the increase could be mainly attributed to an increase in both unknown \( (p < .0001) \) and suspect \( (p < .0001) \) tracks according to machine reasoning and an increase in unknown tracks alone \( (p < .0001) \) according to human reasoning. In synopsis, the data show that the smuggling scenarios are more ‘difficult’ than the traditional ones in terms of ambiguous tracks.

Figure 8. The number of pending tracks, average reaction times for the identification process, and number of signals as a function of scenario type
Furthermore, the data shows an increase in pending tracks in the smuggling scenarios ($p < .001$) (i.e. more pending tracks per time unit) indicating that the human required more time to provide an initial identity in the smuggling scenarios as compared to traditional scenario type (see Figure 8 top). Furthermore, the averaging response times over scenarios disseminates that the response times to identification signals in the traditional scenarios is lower ($F(1,12) = 5.4187, p < .05$) as compared with the smuggle scenarios (see Figure 8 middle). In addition, the number of signals per time unit was significantly higher in the smuggling scenarios ($F(1, 154) = 18.081, p < .0001$) as compared to the traditional scenario indicating that an increased number of tracks are awaiting attention of the human operator (see Figure 8 bottom). These signals requiring attention indicate work to be done and such an increase convey that the human operator requires more effort to get the work done. To summarize, the data reveals three indicators of declined performance in the more difficult scenarios.

A time analysis (see Figure 9 top) reveals an effect of time and scenario type on ambiguous track class ($F(26, 306) = 1.5485, p < .05$). Fisher’s test reveals that the difference manifests itself mainly in the beginning of the scenarios as the first four times slots of the traditional scenarios show significantly less ambiguous tracks than the smuggle scenarios (all $p < .001$).

Figure 9. Top: The number of signals per time unit split by timeslot and scenario type reveal significant different in the first three time slots. Bottom: The number of difficult tracks split by timeslot and scenario type reveals that the scenarios differed mainly in the beginning.

Applying the same time analysis to the number of signals shows for the first three time slots significant less signals in the traditional scenario as compared to the smuggle scenario (all $p < .01$, see Figure 9 bottom). An increasing number of signals represents the fact that the human view and the machine view are increasing in skew as well. This correlation between
number of difficult tracks and number of signals is upheld during the remains of the scenarios. As a larger number of signals requires more attention of the human this can be interpreted as work to be done. Combining the difference in ambiguous tracks with the difference in signals leads to the conclusion that we are not only able to observe overall differences in scenarios or performance, but also to pinpoint those differences in time.

With respect to the effect of scenario type on subjective workload, contrary to expectation we failed to find any subjective workload effects \( F(1,154) = 1.0288, p = .31 \).

Analysis with respect to different behavior of subject rank shows an effect \( F(1, 154) = 4.1954, p < .05 \) of subject rank on signals per time unit in that the warfare officer has more signals per time unit when compared to the assistant. Furthermore, detailed analysis reveals an additional effect of scenario type on subject rank behavior \( F(1, 154) = 5.0065, p < .05 \). Post hoc (Tukey) analysis learns that subject rank behavior manifests mainly in the more difficult scenarios \( p < .001 \) in that the warfare officer has more tracks per time unit (see Figure 10).

9.8 Conclusions

Although the experiment was not designed specifically to validate variation of the CTL variables on workload, we were nevertheless able to determine that:

1. the smuggle scenarios contain more ‘theoretically difficult’ (ambiguous) tracks;
2. the more ‘difficult’ scenarios in terms of ambiguous tracks had a lower performance in terms of pending contacts, response times, and signals awaiting attention and thus were experienced as more difficult by the subjects;
3. the difference in the two types of scenario manifested strongest at the start of the scenarios which correlated nicely with an increase in signals that conveyed the fact that the human operator required more effort to get the work done;
4. there was no effect of scenario type on the subjective workload; and
5. there existed a difference in behavior dependent on subject rank that discriminated in the more ‘difficult’ scenarios.

Taking these five statements into account, we conclude that two of the three hypotheses are clearly confirmed. First of all, although they were not expressly designed as such, the smuggle scenarios are more difficult in terms of ambiguous tracks. Second, there was a clear correlation with the theoretical difficulty of a scenario or situation in terms of ambiguous tracks and the performance of the subjects. We therefore conclude that describing scenarios
in terms of ‘difficult’ tracks is feasible. Such an environmental description in terms of expected workload can be very useful for distilling causes of a performance decrease or an increase in workload. This knowledge, in its turn, benefits the determination of the optimal aiding strategy (i.e. optimizing the what to support question of Endsley, 1996).

In addition, we are not only able to indicate overall differences in scenarios or performance, but also to locate those differences in time due to a combination of the difference in difficult tracks with the difference in signals. This knowledge aids in determining when support is opportune.

The failure to measure subjective-workload effects due to scenario type clearly rejects hypothesis 2 (statement 4). We were, however, able to show a clear performance variation due to a variation in scenario type (statement 2), indicating a larger objective workload in terms of ‘things to do’. The data show performance effects in terms of the number of pending tracks awaiting identification by the user, the number of signals indicating work to be done (objects to inspect and identify), and reaction times to these signals.

Failing to find a subjective workload effect but finding performance effects is attributed by us to a restricted focus of attention by the subjects on the more important objects with at the same time an acceptance of a larger risk due to the diminished attention to the other objects. Humans are capable of maintaining their workload at a ‘comfortable’ level (i.e., level three on a five-point scale) while accepting an increased risk due to not finishing tasks. This notion matches the adaptive-operator theory of Veltman & Jansen (2004) that argues that human operators can utilize two strategies to cope with increased demands. The first strategy involves investing more mental effort and the second involves reducing the task goals. For this second strategy Veltman & Jansen state that ‘operators will slow down the task execution, will skip less relevant tasks, or accept good instead of perfect performance’. As an example, the Tenerife air crash in 1977 was partly attributed to the acceptance of increased risk (Weick, 1993). In our case the subjects seemed to accept the larger risk of not identifying all contacts by limiting their attention to a smaller area around the ship in order to maintain their mental effort at a reasonable level. This is an applied strategy for operational situations where watches take eight hours and it is not known how long an effort must be maintained and it appears the same strategy was followed during the experiments. As a matter of fact, one of the subjects stated as much in saying that ‘his workload should have been five for much of the time as he did not get all the work done’ (i.e., he did not identify all tracks).

Adaptive aiding strategies should consequently be cautious using human indicators of workload only and include at least some performance measures.

The third hypothesis stated that warfare officers show a different behavior in dealing with the tracks compared to warfare assistants. The data indicate evidence in support of this hypothesis (statement 5). Different capabilities, experience, and function show different behavior in that the warfare officer allowed more signals per time unit as compared to the warfare assistants. We argue that this is due to the rank and function-dependent training and background. The assistant warfare officer is trained to construct a complete maritime picture while the warfare officer is supposed to deal with the difficult cases that (potentially) represent a threat. The fact that warfare officers allowed more signals per time unit in the more difficult scenarios indicates that they focused on the more difficult cases and tended to leave the easy cases for the assistant (not present in these single-user experiments). This behavior did not manifest as strongly in the traditional scenarios as these are easier,
resulting in an improved performance in that the warfare officers were not required to focus on the difficult cases alone.

Finding such differences that were not taken into account initially (see section 7.1, 7.2, and 7.3) shows that studies like these are very useful in order to improve cognitive modeling. We conclude that the hybrid model is capable of triggering adaptive automation for a number of reasons. First, the operator performance model optimizes the timing of support and second the model of operator cognition indicates how much work is to be expected in the short term. Third, the model helps optimizing the type of aiding based on the cause of the increased workload.

10. Summary

This chapter took as a starting point that the human-machine ensemble could be regarded as an adaptive controller where both the environment and human cognition vary, the latter due to environmental and situational demands. We have described a human operator performance model and a model of human operator cognition that describe the variability of the human controller as a function of the situation. In addition, we have provided an empirical foundation for the utilization of the combined models. The data from two different experiments show either a change in subjective workload or a performance effect that correlate nicely when the environment or situation is varied.

Both the operator performance model and the model of operator cognition therefore show potential to be used as triggering mechanisms for adaptive automation, or as a measure of a human operator as a slowly changing parameter in an adaptive control system.

11. Acknowledgement

The work described in this chapter was supported by the MOD-NL DR&D organization under programs ‘human supervision and advanced task execution’ (V055) and ‘human system task integration’ (V206). We like to thank Harmen Lafeber, a master student under the auspice of the University of Utrecht, whose master thesis work contributed to experiment I. In addition a number of colleagues from TNO Defence, Security and Safety are thanked for their contribution to the development of the prototype platform, the experiment, and reflective ideas: Jasper Lindenberg, Jan van Delft, Bert Bierman, Louwrens Prins, Rob van der Meer and Kees Houttuin were part of our team in these programs. The eight officers of the Royal Netherlands Navy are thanked for their contributions to experiment II. Finally, we like to thank Willem Treurniet, Erik Willemesen, and Jasper Lindenberg for their useful comments during the writing of this chapter.

12. References


