Towards a model of human monitoring performance

Carmen Bruder, Catrin Hasse and Dietrich Grasshoff

Keywords: automation, human monitoring behaviour, personnel selection, eye tracking, future requirements

Abstract

Corresponding to prior research on the future of aviation, operators will have to monitor highly automated systems appropriately. Future personnel selection tests must address this skill. Eye tracking allows monitoring behaviour to be measured directly. In a previous study, we identified time sensitive eye-tracking parameters for selecting operators monitoring appropriately. Now, the question arises of how to capture human monitoring in the context of other human abilities. Data from an experiment with 90 job candidates were reanalysed and differentiated monitoring performance from other abilities and personality traits. The results show that monitoring performance explains independent portions of variance. Furthermore, monitoring performance has some small, significant relations to certain ability and personality factors. In conclusion, it is useful and important to measure monitoring performance in a separate test.

Zusammenfassung


Introduction

According to the Single European Sky ATM Research (SESAR) Program, a high-performance Air Traffic Management (ATM) infrastructure will exist in Europe in the future. The DLR-project ‘Aviator’ deals with changes that will concern pilots and air traffic controllers in the future with the objective of adapting selection profiles to suit future ability requirements. Workshops were conducted with experienced pilots and air traffic controllers in order to gather their expectations about their future tasks, roles and responsibilities. They indicate that monitoring and teamwork in a highly automated workplace pose challenges to future operators in aviation (Bruder, Jörn & Eißfeldt, 2008). In the future, operators will have to monitor the dynamic processes of automated systems. The increase in automation requires operators monitoring appropriately (OMA), who are able to control the system manually when necessary. This raises the question of which parameters of monitoring behaviour help identify OMA who fit future human-machine systems in aviation.

Identifying operators monitoring appropriately

According to models of adequate and efficient monitoring behaviour (Niessen & Eyferth, 2001; Wickens et al., 2001) as well as differences between experts and novices (Underwood et
al. 2003), it can be stated that OMA demonstrate target-oriented attention allocation both in
general and during monitoring phases, i.e. orientation phase, anticipation phase, operation
phase, and debriefing phase. Whereas the first assumption requires the operator to adapt atten-
tion allocation to the specific requirements of a given situation in general, the second assump-
tion focuses on allocation of attention in phases.

In a previous experiment we identified time sensitive eye-tracking parameters which are
suitable to serve as basis for identifying OMA in future selections (Hasse, Bruder, Grasshoff &
Eißfeldt, 2009a; 2009b). It is assumed that “good monitoring” is associated with accurate ma-
nual control of the system if the automation fails. Therefore performance when actively con-
trolling is used as a criterion for predicting the validity of the eye-tracking parameters. As ex-
pected, there is a significant correlation between participant’s appropriate fixations and manual
performance (r= -.214; p=.045). Besides this substantial effect, the question arises of how to
capture the construct of human monitoring in the context of other human abilities, especially
those abilities required in aviation.

**Identifying moderators of monitoring behaviour**

Recent research suggests that the optimal monitoring and handling of automated systems de-

dpends on certain operator characteristics, such as personality traits and abilities.

The importance of complacency in human use of automation has been clearly established
(Singh, Molloy & Parasuraman, 1993). Complacency addresses the risk of an inappropriate
level of trust placed in the automation by a human operator (Lee & See, 2004). Such excessive
trust can lead to over-reliance on the automated system, without recognising its limitations and
the possibility of automation failure. Several studies have demonstrated that particularly highly
and consistently reliable systems give rise to complacency effects (Singh, Molloy & Parasura-
man, 1993). The more reliably an automated system works, the higher the potential for com-
placency (Prinzel, DeVries, Freeman & Mikulka, 2001). We assumed these attitudes to be also
connected with monitoring performance. From this we hypothesize that complacency potential
improves the predictive power of monitoring behaviour.

Ability tests are used to measure a broad range of knowledge a person brings to the job situ-
ation (Rathje, 2002). The pre-selection of the DLR’s Department of Aviation and Space Psy-
chology consists of a battery of tests covering all important abilities such as memory capacity,
spatial orientation, concentration, attention, numerical abilities, personality as well as some
knowledge-based aspects like English language competency and mechanical comprehension
(Eißfeldt & Deuchert, 2002). Current ability tests for airline pilots and air traffic controllers
measure the ability requirements which are needed to be successful under current training and
job situations (Lorenz, Pecena & Eißfeldt, 1995). This raises the question of whether one’s
monitoring performance could be predicted by these ability tests as well. We assume that the
DLR’s tests concerning attention and concentration are related to monitoring behaviour.

**Method**

We have developed a simulation tool that allows the assessment of monitoring performance
and its relation to other personality and ability factors.

**Simulation tool**

The simulation tool represents a traffic flow simulation (see Figure 1). The traffic flow simul-
ation can be controlled either automatically or manually by a human operator. During the auto-
nomatic phase, the system works fully automatically and the reliability is perfect. During manual
phase, a human operator controls the dynamic traffic manually. This allows performance data
to be collected separately for both types of tasks.
The task of both the automated and human operator control settings is to bring all actual values into agreement with target values (for further information, see Hasse, Bruder, Grasshoff & Eißfeldt, 2009b). Four scenarios, each with a different degree of difficulty, were developed by varying the complexity and dynamics of the automatic system.

Measurements
Monitoring behaviour was measured by recording eye movements (for further information, see Eißfeldt et al., 2009). A test subject’s performance was measured during the manual phase of each scenario. Complacency potential was surveyed using the questionnaire CaP (Complacency as Potential; Feuerberg, Bahner & Manzey, 2005). It consists of four scales reflecting technology-related attitudes (e.g. trust in technology, scepticism towards technology, enthusiasm about technology, locus of control using technology), and four scales concerning personality traits (unconcern/optimism, tolerance of ambiguity, self-efficacy, conscientiousness). Since all the participants were job candidates, ability test measurements, such as memory capacity, attention and concentration, were carried out anyway.

Procedure
Participants were first given the instructions for the experiment as well as a questionnaire measuring their complacency potential. They were informed they would work on four scenarios, each consisting of two phases: first an automated phase and then a manual phase. For the automatic phase of each scenario, participants were instructed to monitor the automated processes with the objective of understanding the rule-based dynamics of the given scenario. For the manual control phase, participants were instructed to manually control the same system they had seen during the automated phase. Eye movement parameters were recorded by the Eye Gaze Analysis System. After a calibration phase (15 s), they were presented with the four scenarios, each with a duration of five minutes. The four scenarios were presented in a fixed order for every subject, beginning with the easiest (Scenario 1) and finishing with the most complex (Scenario 4). After each scenario, subjects evaluated their difficulty and also reported the strategies they identified during the automated phase.
Participants

The experiment was conducted with a sample of 90 candidates for DFS (Deutsche Flugsicherung GmbH) and DLH (Deutsche Lufthansa AG) ranging in age from 17 to 26 years. 82% were male. Experiments were conducted in conjunction with the regular selection process at the German Aerospace Center’s Department of Space and Aviation Psychology. Candidates received 20 € for participating and were assured that their performance in the experiment would not affect their selection results.

Results

First analyses indicate that scenario one and four did not differentiate sufficiently between participants. This is due to ceiling and floor effects. In contrast, scenario two and three show an optimal degree of difficulty. Therefore, results of these two scenarios are reported here.

The effects of complacency potential were analysed. To get an overview, the thirty participants with the best monitoring behaviour were compared to the thirty with the worst. The groups were identified according to their relative fixation counts on relevant AOIs during all operating phases. Contrasts were calculated between the extreme groups. No effect was found for technology-related complacency but the groups differed in the scales conscientiousness (t(84)= 3.39, p<.01), and tolerance of ambiguity (t(86)= -2.97, p<.01). Operators with good monitoring behaviour were significantly more conscientious and less tolerant towards ambiguity. These two significant parameters were involved in further analysis (see Table 1). Additional analysis using a stepwise multiple regression analysis found that tolerance of ambiguity directly influences the performance in the MonT simulation (Mahlfeld, Hasse, Grasshoff & Bruder, 2011). The lower the operator’s value in this scale the better their performance. Conscientiousness has a moderating impact. The relationship between good monitoring behaviour and good performance increases at lower levels of conscientiousness.

Tab. 1: Results of multiple regression analysis with interactions

<table>
<thead>
<tr>
<th>Prediction of Performance</th>
<th>b</th>
<th>t</th>
<th>r</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tolerance of Ambiguity</td>
<td>0.249</td>
<td>0.062*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tolerance of Ambiguity</td>
<td>0.176</td>
<td>2.696**</td>
<td>0.376</td>
<td>0.141**</td>
</tr>
<tr>
<td>Interaction Monitoring</td>
<td>0.163</td>
<td>2.686**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour × Conscientiousness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In addition, the effect of the subjects’ performance in the ability tests was analysed. Four tests were included in this analysis: a visual mental arithmetic test (WSB, attention & concentration), a mental concentration test (KBT, attention & concentration), a visual memory test (MEK, memory capacity) and an acoustic clearance test (CLE, memory capacity). Planned comparisons were performed between these ability tests and relative fixation counts on relevant AOIs during all operating phases. Ability tests of attention and concentration show consistent and significant correlations with monitoring behaviour. No substantial correlations between memory capacity tests and monitoring behaviour were found (see Table 2). Additionally, comparisons between ability test performance and manual control performance were calculated; neither substantial nor consistent correlations were found.
Tab. 2: Correlation analysis between ability test performance and monitoring behaviour.

<table>
<thead>
<tr>
<th>Monitoring behaviour</th>
<th>Attention &amp; concentration</th>
<th>Memory capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WSB</td>
<td>KBT</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>.36*</td>
<td>.25</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>.54***</td>
<td>.42**</td>
</tr>
</tbody>
</table>

* p<.05, ** p<.01, ***p<.001

Discussion & further research

To get a deeper understanding of operational monitoring performance, monitoring performance was related to other relevant parameters, such as candidates’ complacency potential, ability test performance and personality traits. Summing up the results, monitoring performance explains variance independently. Furthermore, monitoring performance has a small, significant relationship with personality traits such as conscientiousness and tolerance of ambiguity, and ability tests such as attention and concentration. Consequently, the effect of personality traits should be taken into account. The relationship between monitoring behaviour and other factors are illustrated by a preliminary integrative model (see Figure 2).

The results show that ability testing using dynamic simulation based on eye movements is innovative and enables new approaches for assessing selection profiles. In this regard, the monitoring test (MonT) is introduced as an appropriate tool for investigating human performance in future ATM scenarios.

As part of the follow-up project Aviator II, we continue to test human monitoring performance. Additional studies are planned to validate our model of adequate monitoring behaviour with experienced controllers and pilots. We plan to replicate this study using the identified monitoring parameters to confirm the effect of personality traits and ability test performance on monitoring behaviour and performance. Furthermore, a team version of MonT is planned, which will enable the assessment of team monitoring performance.

Literature


