Human Performance Consequences of Automated Decision Aids in States of Sleep Loss

Juliane Reichenbach, **Linda Onnasch**, and **Dietrich Manzey,** Berlin Institute of Technology, Berlin, Germany

Objective: The authors investigated how human performance consequences of automated decision aids are affected by the degree of automation and the operator's functional state.

Background: As research has shown, decision aids may not only improve performance but also lead to new sorts of risks. Whereas knowledge exists about the impact of system characteristics (e.g., reliability) on human performance, little is known about how these performance consequences are moderated by the functional state of operators.

Method: Participants performed a simulated supervisory process control task with one of two decision aids providing support for fault identification and management. One session took place during the day, and another one took place during the night after a prolonged waking phase of more than 20 hr.

Results: Results showed that decision aids can support humans effectively in maintaining high levels of performance, even in states of sleep loss, with more highly automated aids being more effective than less automated ones. Furthermore, participants suffering from sleep loss were found to be more careful in interaction with the aids, that is, less prone to effects of complacency and automation bias. However, cost effects arose that included a decline in secondary-task performance and an increased risk of return-to-manual performance decrements.

Conclusion: Automation support can help protect performance after a period of extended wakefulness. In addition, operators suffering from sleep loss seem to compensate for their impaired functional state by reallocating resources and showing a more attentive behavior toward possible automation failures.

Application: Results of this research can inform the design of automation, especially decision aids.

Keywords: human-automation interaction, automation bias, complacency, sleep loss

Address correspondence to Dietrich Manzey, Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, Fachgebiet Arbeits-, Ingenieur- und Organisationspsychologie, Marchstrasse 12, Sekr. F 7 D-10587 Berlin, Germany; e-mail: dietrich.manzey@tu-berlin.de.

HUMAN FACTORS
Vol. 53, No. 6, December 2011, pp. 717-728
DOI:10.1177/0018720811418222

INTRODUCTION

When providing humans with automated decision aids, system designers usually intend to provide some kind of cognitive support that is expected to yield both an improvement of the system's overall reliability and performance as well as a reduction of workload. However, it has been shown that the benefits of automation can be offset by unwanted performance consequences, including an overreliance in the automated functions, a loss of situation awareness, or a loss of manual skills needed to perform the automated functions in case of automation failure (Parasuraman, Sheridan, & Wickens, 2000). To guide research on automation-induced performance effects, several framework models have been proposed (e.g., Endsley & Kaber, 1999; Parasuraman et al., 2000). Common to these models is the assumption that automation does not represent an all-or-none phenomenon but that performance consequences of automation will depend directly on what functions are automated to what extent.

During the past two decades, several studies have been conducted to better understand the specific issues of human-automation interaction. Most of this research has addressed the possible impact of system-related characteristics, such as reliability (e.g., Bailey & Scerbo, 2007; Parasuraman, Molloy, & Singh, 1993), or the kind of function allocation between human and machine (e.g., Endsley & Kaber, 1999, Manzey, Reichenbach, & Onnasch, 2008; Wickens, Huyiang, Santamaria, Sebok, & Sarter, 2010). Other studies have investigated effects of workload (Parasuraman et al., 1993) or individual differences in terms of personality (Szalma & Taylor, 2011).

However, what has gained surprisingly less attention, thus far, is to what extent performance effects of automation are moderated by the functional state of operators. According to Hockey (2003) the functional state can be defined "as the variable capacity of the operator for effective task performance in response to task and environmental demands, and under the constraints imposed by cognitive and physiological processes that control and energise behaviour" (p. 3). Most relevant in this respect are effects of stress and fatigue. Although these issues have repeatedly been discussed in the context of adaptive automation (e.g., Byrne & Parasuraman, 1996), only few studies have addressed such effects within the context of nonadaptive systems.

For example, one set of studies has investigated the effects of 1 night of sleep loss on fault management in a supervisory control task (Hockey, Wastell, & Sauer, 1998; Sauer, Wastell, Hockey, & Earle, 2003). The results of these studies suggest that operators might be able to protect their performance in what they perceive to be their primary task (i.e., fault diagnosis and management). However, this ability seems to be possible only at the expense of increased effort, the use of less demanding strategies of information sampling, and a raised level of attentional selectivity, reflected in declined performance in secondary tasks. Theoretically, this pattern of effects is explained within the framework of the compensatory control model (Hockey, 1997). According to this model, stress and fatigue lead to adaptive regulatory processes that aim at protecting high-priority performance goals at the expense of low-level goals.

The present study extends this line of research by investigating how the interaction with automated aids differing in degree of automation (Wickens et al., 2010) is affected by an impairment of the functional state of operators. The model used for this research includes effects of sleep loss induced by a waking phase of 20 hr or more on the interaction with automated aids in a supervisory process control task. Two kinds of aids are compared. Whereas the first aid (information analysis [IA] support) provides an automatically generated diagnosis only for a given system fault and leaves planning and implementation of all necessary actions to the operator, the second aid (action implementation [AI] support) also performs all steps of fault management autonomously if confirmed by the operator. These two aids were chosen to generate two kinds of automation support that clearly differ with respect to the degree of automation (DOA; Wickens et al., 2010). It is expected that providing the more highly automated aid will make it easier for operators to maintain primary task performance when they are sleep deprived.

However, the most interesting question is to what extent an impaired functional state will promote a misuse of decision aids (Parasuraman & Riley, 1997). On a behavioral level, misuse of automation is reflected in insufficient monitoring of automated functions, referred to as "automation induced complacency" (Parasuraman et al., 1993; Parasuraman & Manzey, 2010). Originally, complacency has been identified as an issue of supervisory control tasks. However, similar effects seem to emerge in interaction with decision aids as well. For example, Mosier and Skitka (1996) have argued that people may use an automated aid "as a heuristic replacement for vigilant information seeking and processing" (p. 205), a phenomenon that they refer to as automation bias.

One sort of automation bias involves operators' following a recommendation of an automated aid even though this recommendation is wrong. According to Skitka, Mosier, and Burdick (1999), such "commission errors can be the result of not seeking out confirmatory or disconfirmatory information or discounting other sources of information in the presence of computer-generated cues" (p. 993). Whereas the latter alternative reflects a typical bias in decision making, the former alternative seems to reflect a decision bias that directly corresponds to complacency in automation monitoring (Parasuraman & Manzey, 2010).

Clear predictions of how sleep loss affects complacency and automation bias are difficult to derive. On the one hand, research from supervisory control tasks suggests that operators suffering from sleep loss tend to apply less cognitively demanding performance strategies and to reduce their information sampling while monitoring the system (Hockey et al., 1998). If similar effects emerge in interaction with decision aids, one might expect that sleep-deprived operators start to rely more on their automated aids, that is, reduce the effort to verify the automated recommendations. On the other hand, these operators might be particularly concerned

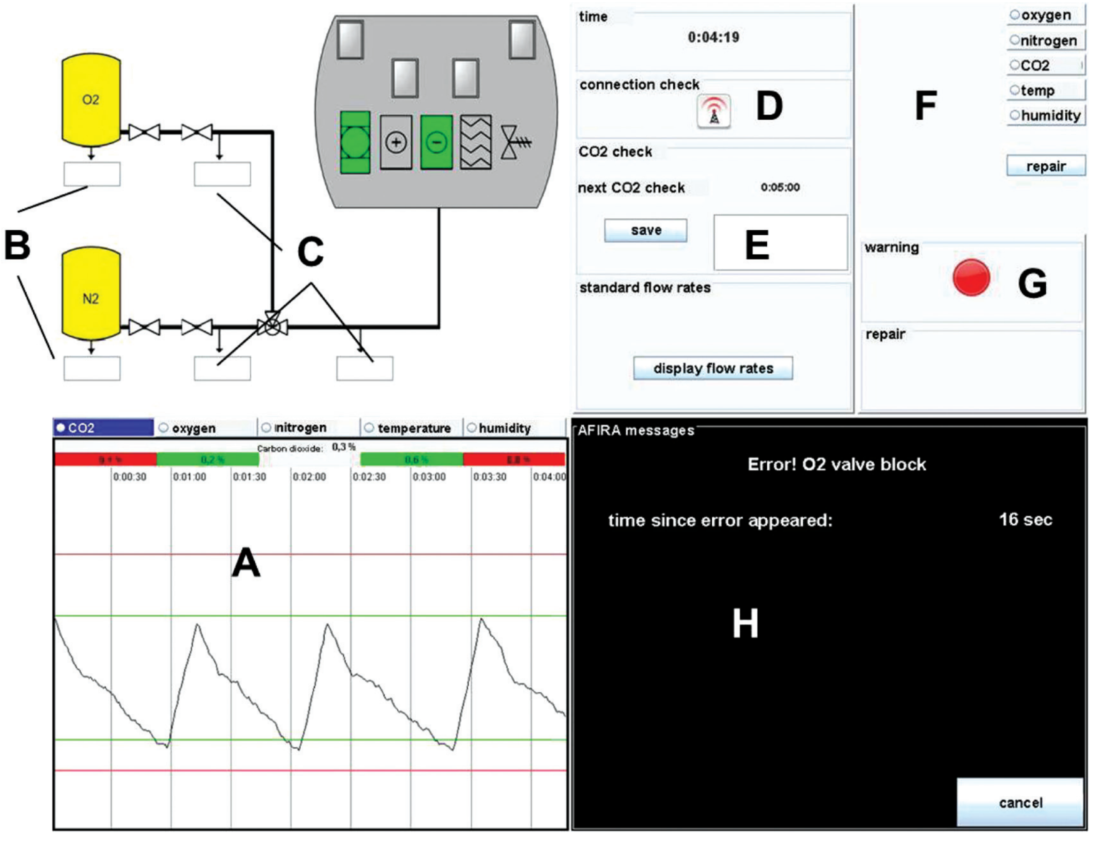

Figure 1. Interface of AutoCAMS 2.0. The figure shows the system with active "information analysis" (IA) support: (A) history graphs, (B) tank level readings, (C) valve-flow readings, (D) icon of probe reaction time task, (E) window for prospective memory secondary task (entry of CO_2 readings), (F) menu for manual control and repair orders, (G) master alarm, and (H) assistance system (Automated Fault Identification and Recovery Agent).

to protect their performance by avoiding errors in interaction with an automated aid and interact more carefully with the system as a consequence. In this case, they might even increase their information sampling and cross-check activities, that is, show a lesser degree of complacency and automation bias. Such effects would be in line with recent results reported by Chaumet et al. (2009) and Killgore (2007), which suggest that sleep deprivation up to 36 hr considerably reduces risk-taking propensity.

METHOD

Participants

For this study, 32 engineering students ranging in age from 19 to 32 years ($M = 24.9$) participated in the study. On the basis of results from a Morningness-Eveningness Questionnaire (Griefahn, Künemund, Bröde, & Mehnert, 2001), extreme evening types were excluded from the experiment. Participants were paid €150 for completing the study.

Apparatus: AutoCAMS 2.0

A "microworld" simulation of a supervisory process control task was used for the experiment (AutoCAMS 2.0; Manzey, Bleil, et al., 2008) that simulated an autonomously running life-support system consisting of five subsystems critical for the maintenance of atmospheric conditions (e.g., oxygen, pressure, carbon dioxide). During nominal operation, all parameters are automatically kept within target range. However, because of malfunctions in the system (e.g., leak of a valve), parameters can go out of range. The interface of AutoCAMS 2.0 is shown in Figure 1.

The primary task of the operator involves supervisory control of the five subsystems, including diagnosis and management of system faults. Whenever a fault is detected in the system, a master alarm turns on, and a time counter starts displaying how much time has elapsed since the occurrence of the fault. To have the malfunction fixed, its specific cause has to be identified and an appropriate repair order has to be selected from a maintenance menu. The repair itself takes 60 s. During this time, the operator is required to control the affected subsystem manually. If the sent repair order is correct, the warning signal turns green and all subsystems run autonomously again. In case of a wrong repair order, the warning light stays red and manual control is required until a correct repair is initiated and completed.

Dependent on the specific condition, participants have to perform fault diagnosis and management manually or with the support of one of two different versions of an automated aid (Automated Fault Identification and Recovery Agent [AFIRA]). In case of IA support, the master alarm is accompanied by a message providing a specific diagnosis for the given fault. Yet action planning and implementation are left to the operator. In case of AI support, AFIRA also implements all necessary steps autonomously if confirmed by the participant. To identify faults in the manual condition or to verify proposed diagnoses in conditions with automation support, operators have independent access to all important parameters. These include relevant system parameters and a history graph for each of the five subsystems. However, this information is not always visible but has to be activated for a 10-s view by mouse click on the tank, flow meter, or history graph.

In addition to the primary task, two secondary tasks have to be performed. The first is a prospective memory task in which participants are required to record the CO_2 values every 60 s. The other is a probe reaction time task. Participants have to click on a "communication link" icon as fast as possible to confirm a proper connection with the cabin. This icon appears in random intervals roughly twice per minute.

Design

The study used a 2 (day vs. night session) × 5 (block) × 2 (DOA) design with DOA (IA support vs. AI support) defined as between-subjects factor and session (day vs. night) and block defined as within-subjects factors. Participants were randomly assigned to the two DOA groups. Half of the participants (n = 16) worked with IA and the other half (n = 16) with AI support. The sequence of day and night session was balanced within each experimental group. The five blocks per session differed with respect to whether automation support was available. Whereas the participants had to perform fault identification and management manually during Blocks 1 and 5, they were supported by the automated aid in Blocks 2 through 4. During the first session (day for half of the participants, night for the other half), six different faults occurred in each block, which were all correctly indicated and diagnosed by the automated aid.

The second session replicated the basic structure of the first one. However, during this session, an additional, seventh fault occurred at the end of Block 4, for which AFIRA provided a wrong diagnosis. This failure of AFIRA was implemented to simulate a "first automation failure effect." Because the failure always occurred during the second experimental session, those participants who started the experiment with the day session experienced this failure at night. The other half of the participants experienced it during the day session. This feature of the study design is illustrated in Figure 2.

Dependent Measures

Primary task performance. Performance measures calculated for each block included (a) percentage of correct diagnoses, that is, the percentage of the faults for which the first repair order sent was correct; (b) fault identification time (FIT), that is, mean time (in seconds) needed to issue a correct repair order; and (c) out-of-target error (OTE), that is, duration (in seconds) that the most critical system parameter (oxygen) was out of target range while a fault was present in the system.

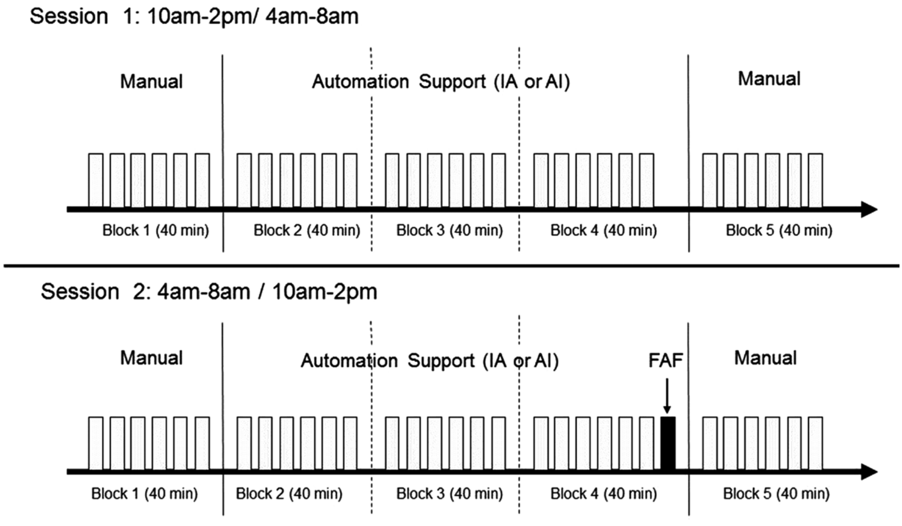

Figure 2. Study design for the two degree-of-automation (DOA) groups. Shown are the sequence and internal structure of experimental blocks during the first (upper panel) and second (lower panel) experimental session. The occurrence of the first automation failure is marked by "FAF." Half of the participants of each DOA group (n =16) had their first session scheduled during the day (10 a.m. to 2 p.m.) and the second session during the night (4 a.m. to 8 a.m.). The other half performed the experimental sessions in the inverse sequence.

Secondary task performance. Performance in the two secondary tasks was assessed by (a) the mean time (in milliseconds) needed to respond to the connection check prompted and (b) the number of timely entries of CO_2 levels (prospective memory task). Only responses during phases when the participant had to deal with a system fault were considered.

Automation verification and automation bias. Measures used to assess the level of complacency included (a) automation verification time (AVT) and (b) automation verification information sampling (AVIS). AVT was defined as the time interval (in seconds) from the appearance of the master warning until sending of a first repair order, regardless whether this order was correct. AVIS was defined as the percentage of system parameters accessed necessary to completely verify a given diagnosis provided by AFIRA. Only parameters accessed between the occurrence of the master warning and sending of the first repair order were considered for this measure. Automation bias was analyzed by the percentage of participants committing a commission error, defined as percentage of participants who followed the diagnosis of the aid for Fault 7 in the last experimental block although it was wrong.

Subjective workload and sleepiness. Subjective sleepiness was assessed by the Stanford Sleepiness Scale (SSS; Hoddes, Dement, & Zarcone, 1972). Performance indicators of sleepiness were derived from the Psychomotor Vigilance Task (PVT), which represents a short-term simple reaction-time task and has particularly been developed to evaluate sleepiness-induced performance impairments (Dinges & Powell, 1985). Measures included the overall mean of reaction times and the mean of the 10% slowest reaction times. Subjective workload was assessed by the NASA Task Load Index (Hart & Staveland, 1988).

Procedure

The experiment consisted of two practice sessions and two experimental sessions across

4 days. The first practice session lasted 4 hr and included familiarization with the AutoCAMS system. Participants were introduced to the different subsystems and were trained to manually identify and manage eight possible faults. On the 2nd day, all participants had to perform a 45-min test trial that served to test their acquired performance skills according to a predefined criterion. All participants passed this test successfully.

On the 3rd day, the first experimental session took place. For half of the participants, this session was scheduled during the day (10 a.m. to 2 p.m.); for the other half, it was scheduled after more than 20 hr of continuous wakefulness during the night at the nadir of the circadian system (4 a.m. to 8 a.m.). Before the session started, participants were assigned to one of the two experimental groups (IA vs. AI support) and were familiarized with using "their" aid. During this trial, all recommendations provided by AFIRA were correct. However, participants were made aware that the aid's reliability was high but not perfect and were explicitly warned to check the proposed diagnoses before initiating a repair. Then the first session of the experiment started, consisting of five blocks of 40 min each. During the first and fifth block, all participants worked without the assistance of AFIRA. During Blocks 2, 3, and 4, they were supported by AFIRA. All blocks were comparable with respect to the set of faults, and the distribution of faults across blocks was the same for both experimental groups. The second experimental session took place 1 week later. It included the same set, sequence, and internal structure of blocks with the only exception that a first automation failure of the AFIRA occurred at the end of Block 4.

Participants were instructed to get up at 8:00 a.m. on experimental days. This was controlled by an actimeter. During both sessions, day and night, sleepiness and workload ratings were collected. The PVT was administered before the first block started. Subjective sleepiness was assessed before the first and after each block. Subjective workload was assessed after each block.

RESULTS

Sleepiness

Effects on subjective sleepiness (SSS) were analyzed by a 2 (session) × 6 (block) × 2 (DOA) ANOVA. A significant main effect of session, $F(1, 30) = 184.21$, $p < .01$, $\eta^2 = .86$; a significant block effect, $F(5, 150) = 10.65$, $p < .01$, $\eta^2 = .26$; and a Session × Block interaction, $F(5, 150) = 2.61$, $p < .03$, $\eta^2 = .08$, were found. On a scale ranging from 1 (*feeling active and vital, alert, wide awake*) to 7 (*almost in reverie, sleep onset soon, lost struggle to remain awake*), the mean level of SSS was higher during the night ($M = 4.67$) than during the day ($M = 2.44$). Furthermore, it increased considerably across blocks during the night session but remained constant on a low level throughout the day session. The PVT results were analyzed by a *t* test. Mean response times were longer after extended wakefulness ($M = 309$ ms) than during the day session ($M = 293$ ms), $t(31) = 4.22$, $p < .01$ $\eta^2 = .37$, and a similar effect emerged for the 10% slowest reactions (446 ms vs. 417 ms), $t(31) = 2.15$, $p < .05$, $\eta^2 = .13$.

Primary Task Performance: Fault Identification and Management

Primary task performance measures were analyzed by a 2 (session) × 5 (block) × 2 (DOA) ANOVA. Because of the operational relevance of performance decrements in impaired functional states, the following report of effects involving the session factor will not be limited to significant effects ($p < .05$); we will also consider effects that approach the conventional level of significance ($.05 < p < .10$).

Percentage of correct fault identifications varied across blocks, $F(4, 120) = 12.87$, $p < .01$, $\eta^2 = .30$. On average, 91.9% and 96.1% of faults were correctly identified in the two manual blocks. With automation support in Blocks 2 through 4, this already high level of performance increased to an almost perfect performance (>98%). No effects of day versus night or DOA were found, but there was a complex three-way interaction, $F(4, 120) = 2{,}70$, $p < .04$, $\eta^2 = .08$, which, however, did not reveal any meaningful pattern of effects.

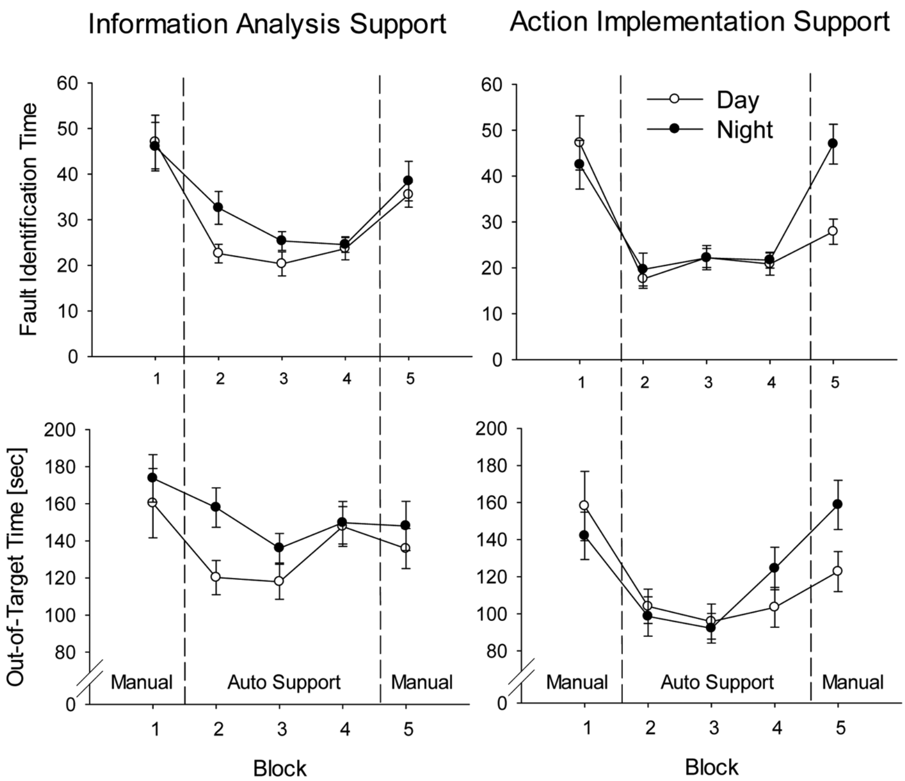

Figure 3. Effects of automation support on primary task performance measures dependent on the functional state of participants (day vs. night session), availability of automation support (blocks), and degree of automation (information analysis vs. action implementation support). Shown are means and standard errors of fault identification times (upper panel) and out-of-target error (lower panel).

Effects for FITs and OTEs are displayed separately for the two DOA conditions in Figure 3. As expected, FIT (upper panel) profited considerably from automation support in Blocks 2 through 4 compared with manual performance in Blocks 1 and 5, $F(4, 120) = 43.59$, $p < .01$, $\eta^2 = .59$. No main effect of DOA emerged, $F(1, 30) = 1.20$, $\eta^2 = .04$, but performance tended to be generally better during the day than during the night session, $F(1, 30) = 3{,}87$, $p < .06$, $\eta^2 = .11$. This pattern of main effects was qualified by a significant Session × Block interaction, $F(4, 120) = 2.97$, $p < .03$, $\eta^2 = .09$, and a Session × Block × DOA interaction, $F(4, 120) = 2.41$, $p = .05$, $\eta^2 = .07$. The sources of the latter interaction can be derived from comparing the time courses of effects shown in the upper left and right panel of Figure 3. Participants working with the less automated aid (i.e., IA support, left panel) initially were less able than the other group to protect performance at night compared with daytime performance (see day-to-night difference in Block 2) but improved over time and did not show any greater difficulties of return-to-manual performance during the night than during the day.

In contrast, participants working with AI support (upper right panel) were able to perfectly maintain performance even in a state of sleep loss, if automation support was available (Blocks 2 through 4). However, they showed considerably greater difficulties of return-to-manual performance during the night than during the day, reflected in the day-to-night performance difference in Block 5. Essentially the same pattern of effects emerged for OTE: block,

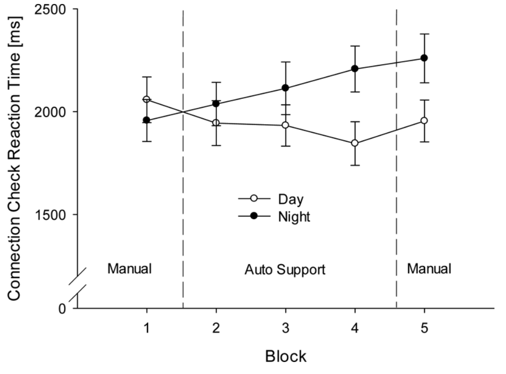

Figure 4. Secondary task response times dependent on functional state of participants (day vs. night session) and availability of automation support (blocks).

$F(4, 120) = 13.47$, $p < .01$, $\eta^2 = .31$; session, $F(1, 30) = 8.55$, $p < .01$, $\eta^2 = .22$; Block × DOA, $F(4, 120) = 2.31$, $p < .07$, $\eta^2 = .07$; Session × Block × DOA, $F(4, 120) = 2.13$, $p < .09$, $\eta^2 = .07$. In addition, a significant main effect of DOA emerged for this measure, $F(1, 30) = 6.22$, $p < .01$, $\eta^2 = .17$. As expected, participants supported by the highly automated aid were more able to keep the oxygen level within the given limits of nominal operation.

Secondary Task Performance: Probe Reaction Times and Prospective Memory

Performance of both secondary tasks was analyzed by a 2 (session) × 5 (block) × 2 (DOA) ANOVA. A significant main effect of session, $F(1, 30) = 8.45$, $p < .01$, $\eta^2 = .22$, moderated by a Session × Block interaction, $F(4, 120) = 3.14$, $p < .02$, $\eta^2 = .10$, was found for the probe reaction times (connection check task; see Figure 4). During the day, an automation benefit was visible with faster response times during automation supported blocks (1,907 ms vs. 2,006 ms). At night, secondary task response times increased over time even in the automation-supported blocks. No effect of DOA emerged in this measure. For the prospective memory task (entry of CO_2 levels), only a significant block effect was found, $F(4, 120) = 21.52$, $\eta^2 = .42$. Participants were more able to make timely entries in the blocks with automation support, compared with manual performance.

Automation Verification

Automation verification time and effort were analyzed by a 2 (session) × 3 (block) × 2 (DOA) ANOVA. Only main effects of session became significant. With respect to both measures, participants were found less complacent when sleepy as compared with being in an alert state. Specifically, they spent significantly more time with automation verification (22.7 vs. 19.6 s), $F(1, 30) = 6.37$, $p < .02$, $\eta^2 = .18$, and sampled more of the system parameters immediately needed to cross-check a given diagnosis (97% vs. 92%), $F(1, 30) = 4.34$, $p < .05$, $\eta^2 = .13$, during the night than during the day session. Neither the main effects of DOA or block nor any of the interaction effects became significant for any of the two measures (all $F < 1.25$).

Automation Bias

We found clear evidence for automation bias by analyzing the responses to Fault 7 in the last block of the second experimental session, whereby the automated aid failed for the first time by providing a wrong diagnosis. The strength of this effect was moderated by the functional state of the participants. Whereas 7 (43.75%) of the 16 participants who experienced this automation failure during the day followed the diagnosis even though it was wrong, only 1 of the 16 participants (6.25%) who experienced the AFIRA failure during the night committed this kind of commission error. This effect was not affected by DOA. A 2 (session) × 2 (DOA) ANOVA revealed only a significant session effect, $F(1, 28) = 6.63$, $p < .02$, $\eta^2 = .19$.

To investigate whether the information sampling behavior of participants who committed the commission error differed from the behavior of participants who detected the automation failure, an additional 2 (group) × 4(block) ANOVA was run, contrasting the extent to which both groups of participants cross-checked the diagnoses of the aids for the 18 system faults in Blocks 2 through 4 for which AFIRA provided a correct diagnosis as well as the critical Fault 7 in Block 4. Only the 16 participants who were confronted with this failure during the day session were considered in this analysis. It revealed a main effect of group, $F(1, 14) = 5.36$, $p < .04$, $\eta^2 = .28$, and a Group × Block interaction, $F(1, 42) = 3.22$,

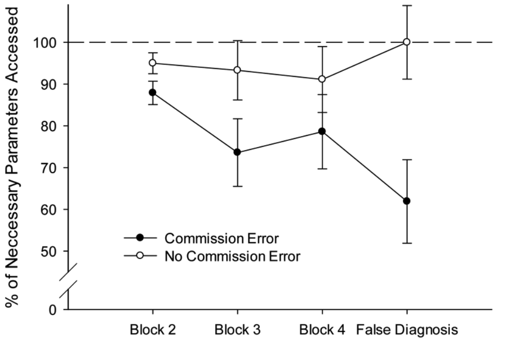

Figure 5. Information sampling behavior of participants who have and have not committed a commission error in case of false diagnosis provided by the aid. Shown are means and standard errors for proportion of system parameters accessed that were needed for automation verification.

$p < .04$, $\eta^2 = .19$. These effects are illustrated in Figure 5.

The 7 participants who committed the commission error checked on average fewer of the system parameters (75.5%) needed for verifying the aid's recommendation than did the 9 participants who did not commit this kind of error (94.9%). Furthermore, participants committing the commission error reduced their automation verification effort over time, whereas the other participants remained careful in cross-checking the aid. For the critical system Fault 7, participants showing the automation bias effect on average cross-checked only 61.5% of the necessary parameters, whereas the other participants detected the wrong diagnosis by fully (i.e., 100%) verifying the aids' diagnosis.

Subjective Workload

Analysis of subjective workload revealed a significant main effect of session, $F(1, 30) = 8.45$, $p < .01$, $\eta^2 = .22$, and a significant block effect, $F(4, 120) = 3.17$, $p < .02$, $\eta^2 = .10$. Participants showed higher workload ratings during the night than during the day. In addition, subjective workload was higher in the first block of each session as compared with all other blocks. Neither the main effect of DOA nor any of the interaction effects became significant, all $F \leq 1.8$. A separate analysis for the Effort subscale, which seems to be most relevant for predictions derived from the compensatory control model, revealed a main effect of session, $F(1, 30) = 20.03$, $p < .01$, $\eta^2 = .40$, and a Session × Block interaction, $F(4, 120) = 2.92$, $p < .03$, $\eta^2 = .09$. As expected, effort was rated higher during the night than during the day and showed a sharp increase after Block 4, when the automation failure occurred.

DISCUSSION

The results provide first insights on how human performance consequences of automated decision aids are moderated by impairments of the functional state of an operator induced by sleep loss. On the basis of the compensatory control model (Hockey, 1997), it was supposed that operators would be able to protect their primary task performance in states of sleep loss by increased effort and/or reallocation of resources between tasks of different priority. It was further assumed that this protection of primary task performance would be easier if higher DOA aids were available for the task. The present results provide at least partial support for this hypothesis.

Participants supported by the high-DOA aid were able to diagnose system faults as rapidly at night, after being awake for more than 20 hr, as during the day (Figure 2, upper right panel, Blocks 2 through 4). This marked a difference from participants supported by the less automated aid, who were less able to protect performance in the primary task in the first block with automated support (Figure 2, upper left panel). A similar pattern of effects was found for the control performance measure. This result suggests that the beneficial effects of automated support in states of sleep loss are to some extent dependent on DOA. Releasing operators from manual tasks in addition to cognitive tasks seems to be a better countermeasure against performance decrements in states of sleep loss than just providing support for cognitive functions.

However, even with the highly automated aid, protection of primary task performance during the night was possible only at the expense of increased perceived effort and a reallocation of attentional resources, reflected in a decrement of secondary task performance. This finding provides direct support for the compensatory control model (Hockey, 1997) and suggests that although highly automated aids increase the efficiency of compensatory processes, they do not suffice to completely prevent such processes. Furthermore, the use of a highly automated aid in states of sleep loss seems to amplify performance decrements if one needs to return to manual performance again. This supports earlier findings by Endsley and Kiris (1995) suggesting that risks of out-of-the-loop unfamiliarity and related issues of return-to-manual performance increase with higher levels of automation.

Probably the most surprising results of the present study relate to the effects of sleep loss on complacency and automation bias. Given results from studies of sleep deprivation on supervisory control processes (Hockey et al., 1998), it was expected that complacency in terms of less demanding automation verification strategies might represent another compensatory control mechanism of operators to unload themselves when they feel sleepy. The present results contradict this hypothesis by revealing the opposite pattern of effects. During the night, participants of both DOA groups were found to invest more time in automation verification and to sample a higher proportion of system parameters than was the case during the day. It seems that participants who have to work in an impaired functional state interact more carefully and attentively with the aid and invest more cognitive resources to verify its recommendation before accepting it.

This finding can also explain the increase in perceived effort and the decrements in secondary task performance reported earlier and might reflect an important aspect of an adaptive compensatory control process that specifically becomes apparent in human-automation interaction. Instead of relying more on automated systems when sleepy, operators appear to show a more attentive behavior to avoid errors and to compensate for the elevated risk attributable to their sleepiness. Such behavioral tendency would be in accordance with recent results that suggest that sleep deprivation generally reduces the propensity to take risks (Chaumet et al., 2009; Killgore, 2007). In addition, it might also reflect the effort of operators to stay actively involved in the work task to counteract the increasing sleepiness. The latter effect would suggest that the finding of increased verification effort might hold true specifically for the use of automated aids in situations when the overall workload of operators is comparatively low and, therefore, the need to actively cope with raised sleepiness is high.

As a consequence of the greater effort of verification, participants of the present study were found to be significantly less prone to automation bias effects in a state of sleepiness. Out of the 16 participants confronted with a first automation failure during the day, almost half ($n = 7$) committed a commission error, and this effect emerged independent of the DOA of the aid. However, after being continuously awake for more than 20 hr, only 1 participant committed this kind of error. This effect emerged even though all participants had worked for the same time with the system when the fault occurred, and none of them knew that such an event had to be expected.

Evidence that this effect is directly linked to differences in automation verification strategies is provided by the results of the microanalyses of information-sampling behavior performed on participants who were confronted with the first automation failure effect during the day. Participants who committed a commission error were found to spend less time on automation verification and to check less necessary parameters than did participants who did not commit such error. This finding replicates results of our earlier research (Bahner, Hueper, & Manzey, 2008; Reichenbach, Onnasch, & Manzey, 2010) and provides more evidence for an integrated model of complacency and automation bias recently proposed by Parasuraman and Manzey (2010). Yet, during the night, automation verification time and information sampling remained at a high level for all participants.

In conclusion, the results of the present study can be taken as a first step toward a better understanding of human performance consequences of automation in impaired functional states. Specifically, they suggest that providing automated aids can represent an effective means for counteracting performance decrements in states of fatigue induced by sleep loss, with more highly automated aids being more effective in supporting routine performance than less automated ones. However, this benefit of high DOA might be associated with issues of return-to-manual performance in case of automation failures. In contrast, concerns that impaired functional states might increase the risks of complacency and automation bias are not supported by the present results.

ACKNOWLEDGMENTS

This research was sponsored by Research Grant MA 5739-1 provided by Deutsche Forschungsgemeinschaft. Thanks are due to Marcus Bleil for software development and Johannes Beck, who helped with the data acquisition during the night sessions.

KEY POINTS

- Automation support can help protect performance in states of fatigue, especially more highly automated aids that release the operator from manual tasks in addition to cognitive tasks.
- However, use of a highly automated aid in states of fatigue seems to amplify performance decrements if one needs to return to manual performance.
- Clear evidence for automation bias was found only for the day session. In a state of sleep loss, operators show a more attentive behavior to avoid errors and to compensate for the elevated risk attributable to their fatigue.

REFERENCES

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues of Ergonomics Science*, *8*, 321–348.

Bahner, J. E., Hueper, A. C., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias, and the impact of training experiences. *International Journal of Human-Computer Interaction*, *66*, 688–699.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, *42*, 249–268.

Chaumet, G., Taillard, J., Sagaspe, P., Pagani, M., Dinges, D. F., Pavy-Le-Traon, A., Bareille, M.-P., Rascol, O., & Philip, P. (2009). Confinement and sleep deprivation effects on propensity to take risks. *Aviation, Space and Environmental Medicine*, *80*, 73–80.

Dinges, D. I., & Powell, J. W. (1985). Microcomputer analysis of performance on a portable, simple visual RT task sustained operations. *Behavioral Research Methods, Instrumentation, and Computers*, *17*, 652–655.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, *37*, 387–394.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness, and workload in dynamic control task. *Ergonomics*, *42*, 462–492.

Griefahn, B., Künemund, C., Bröde, P., & Mehnert, P. (2001). Zur Validität der deutschen Übersetzung des Morningness-Eveningness-Questionnaires von Horne und Östberg [Introduction: Operator functional state in the analysis of complex performance]. *Somnologie*, *5*, 71–80.

Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam, the Netherlands: Elsevier.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*, 73–93.

Hockey, G. R. J. (2003). Introduction: Operator functional state in the analysis of complex task performance. In G .R. J. Hockey, A. W. K. Gaillard, & O. Burov, (Eds.), *Operator functional state: The assessment and prediction of human performance degradation in complex tasks* (pp. 3–7). Amsterdam, the Netherlands: IOS.

Hockey, G. R. J., Wastell, D. G., & Sauer, J. (1998). Effects of sleep deprivation and user interface on complex performance: a multilevel analysis of compensatory control. *Human Factors*, *40*, 233–253.

Hoddes, E., Dement, W. C., & Zarcone, V. (1972). The development and use of the Stanford Sleepiness Scale (SSS). *Psychophysiology*, *10*, 431–436.

Killgore, W. D. S. (2007). Effects of sleep deprivation and morningness-eveningness traits on risk-taking. *Psychological Report*, *100*, 613–626.

Manzey, D., Reichenbach, J., & Onnasch, L. (2008a). Performance consequences of automated aids in process control: The impact of function allocation. *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 297–301). Santa Monica, CA: Human Factors and Engineering Society.

Manzey, D., Bleil, M., Bahner-Heyne, J. E., Klostermann, A., Onnasch, L., Reichenbach, J., & Röttger, S. (2008b). *AutoCAMS 2.0 manual*. Retrieved from http://www.aio.tu-berlin.de/?id=30492

Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 201–220). Mahwah, NJ: Lawrence Erlbaum.

Parasuraman, R., & Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*, 381–410.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation induced "complacency". *International Journal of Aviation Psychology*, *2*, 1–23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*, 230–259.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, *30*, 286–297.

Reichenbach, J., Onnasch, L. & Manzey, D. (2010). Misuse of automation: The impact of system experience on complacency and automation bias in interaction with automated aids. *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 374–378). Santa Monica, CA: Human Factors and Ergonomics Society.

Sauer, J., Wastell, D., Hockey, G. R. J., & Earle, F. (2003). Performance in a complex multiple-task environment during laboratory-based simulation of occasional night work. *Human Factors*, *45*, 657–669.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*, 991–1006.

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: The big five factors of personality. *Journal of Experimental Psychology: Applied*, *17*, 71–96. doi:10.1037/a0024170

Wickens, C. D., Huyiang, L., Santamaria, A., Sebok, A., & Sarter, N. B. (2010). Stages and levels of automation: An integrated analysis. *Proceedings of the 54th Annual Meeting of the Human Factors and Society* (pp. 389–393). Santa Monica, CA: Human Factors and Ergonomics Society.

Juliane Reichenbach is a PhD candidate in the Department of Psychology and Ergonomics of Berlin Institute of Technology. She has obtained a master in psychology in 2004 from University of Regensburg. The current article represents a part of her dissertation work.

Linda Onnasch is a research fellow in the Department of Psychology and Ergonomics of Berlin Institute of Technology, where she has obtained a master in psychology in 2009. She is currently working on a PhD addressing issues of trust in automation.

Dietrich Manzey is a university professor of work, engineering, and organizational psychology at the Institute of Psychology and Ergonomics, Berlin Institute of Technology, Germany. He received his PhD in experimental psychology at the University of Kiel, Germany, in 1988 and his habilitation in psychology at the University of Marburg, Germany, in 1999.

Date received: February 8, 2011
Date accepted: July 3, 2011