



Topics in Cognitive Science (2020) 1–18

© 2020 The Authors. *Topics in Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society

ISSN:1756-8765 online

DOI: 10.1111/tops.12515

This article is part of the topic “Best of Papers from the 17th International Conference on Cognitive Modeling,” Terrence C. Stewart and Christopher Myers (Topic Editors). For a full listing of topic papers, see [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1756-8765/earlyview](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview)

## A Neuroadaptive Cognitive Model for Dealing With Uncertainty in Tracing Pilots’ Cognitive State

Oliver W. Klaproth,<sup>a</sup>  Marc Halbrügge,<sup>b</sup> Laurens R. Krol,<sup>c</sup>  
Christoph Vernaleken,<sup>d</sup> Thorsten O. Zander,<sup>c</sup> Nele Russwinkel<sup>b</sup>

<sup>a</sup>*Airbus Central Research & Technology*

<sup>b</sup>*Chair of Cognitive Modelling in Dynamic Systems, TU Berlin*

<sup>c</sup>*Zander Laboratories B.V.*

<sup>d</sup>*Airbus Defence & Space*

Received 10 January 2020; received in revised form 6 May 2020; accepted 8 June 2020

---

### Abstract

A model-based approach for cognitive assistance is proposed to keep track of pilots’ changing demands in dynamic situations. Based on model-tracing with flight deck interactions and EEG recordings, the model is able to represent individual pilots’ behavior in response to flight deck alerts. As a first application of the concept, an ACT-R cognitive model is created using data from an empirical flight simulator study on neurophysiological signals of missed acoustic alerts. Results show that uncertainty of individual behavior representation can be significantly reduced by combining cognitive modeling with EEG data. Implications for cognitive assistance in aviation are discussed.

**Keywords:** Cognitive modeling; Flight deck alerts; Cognitive assistance; Model-tracing; Neuroadaptive technology

---

---

Correspondence should be sent to Oliver W. Klaproth, Airbus Central Research & Technology, Hein-Sass-Weg 22, 21129 Hamburg, Germany. E-mail: [oliver.klaproth@airbus.com](mailto:oliver.klaproth@airbus.com)

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

Cognitive assistance is about providing the right information at the right time. The quality of support that can be provided depends on what is and can be known about the task environment and the operator's cognitive processes. On the flight deck, environment and task dynamics make it hard to predict pilot behavior and cognitive states as a function of the environment. For such complex tasks, very extensive models would be needed to incorporate all sources of variability for explaining individual performance in a deterministic fashion. Furthermore, individual differences such as experience on the aircraft type can make explaining and predicting individual performance very challenging.

Rather than modeling all potential sources of variability, simpler models are needed to anticipate pilot behavior. Such models need to be able to continuously update information about the dynamics of the environment and task as well as individual differences in real time to anticipate pilots' individual needs for assistance. The implementation of such capabilities for cognitive assistance should draw inspiration from how support according to anticipated individual needs and actions is accomplished in human-human interactions.

### 1.1. Anticipating individual behavior

In cooperative tasks, the ability to act together is crucial, even in simple activities such as washing the dishes or sharing a bicycle lane with other cyclists. According to Sebanz, Bekkering, and Knoblich (2006), joint action can be regarded as any form of social interaction where two or more individuals coordinate their actions in space and time to bring about a change in the environment. A successful joint action depends on the abilities (a) to share representations, (b) to predict actions, and (c) to integrate predicted effects of one's own and others' actions.

The process of joint attention, that is, monitoring another individual's perceptual processes and directing one's own attention to the same objects and events, provides a basic mechanism for sharing representations and achieving "perceptual common ground" (Sebanz et al., 2006). Together with detailed knowledge about the cooperators' task, this allows one to make educated predictions of another's cognitive dynamics and overt actions (Vesper et al., 2016). Then, the observation of actions both helps in validating predictions and understanding others' cognitive states. For example, studies have shown that observation of an action leads to activation of a corresponding representation in the observer's action system (e.g., Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006; Grèzes, Armony, Rowe, & Passingham, 2003).

Cognitive models can be used to provide cognitive assistance on the flight deck with capabilities such as anticipation of actions in cooperative tasks similar to those of humans. Most cognitive models, however, are built to simulate average user behavior under controlled conditions instead of individual performance in complex tasks (Rehling, Lovett, Lebiere, Reder, & Demiral, 2004). Representing individual behavior in naturalistic settings requires dealing with multiple sources of variation such as inter-individual

differences (e.g., architectural and knowledge differences; Taatgen, 1999) and uncontrolled external factors of the situation. For example, when modeling pilot performance in commercial aviation, different levels of experience and changing weather conditions would need to be considered. A cognitive model that is able to keep track of the operational context and an individual user's cognitive dynamics can serve as the basis for cognitive assistance in operations (Zhang, Russwinkel, & Prezenski, 2018).

## 1.2. Cognitive modeling of uncertainty

In naturalistic settings, uncertainty about the individual user's cognitive state needs to be negotiated when making decisions about how and when to support. Uncertainty can be divided into (a) epistemic uncertainty that is caused by a deficiency in knowledge and (b) aleatory uncertainty induced by the inherent variability of events (Kiureghian & Ditlevsen, 2009). While acquiring or providing additional information can decrease the epistemic proportion of uncertain situations, aleatory uncertainty cannot be reduced but needs to be managed carefully instead. In experimental psychology, distinct coping strategies associated with each type have been identified in human decision-making under uncertainty: When faced with epistemic uncertainty, information search or exploration processes are elicited, while aleatory uncertainty is usually met with exploitation behavior, that is, relying on the most probable outcome to maximize reward (Fox & Ülkümen, 2011). This distinction between uncertainties and the associated behavioral responses have both architectural (a) as well as cognitive (b) implications when building models for assistance to individual user behavior.

(a) When trying to represent individual behavior, epistemic uncertainty could be generally reduced by including more information about the user and the world in the model. In real-world or realistic settings, however, environmental dynamics might be hard to capture, and substantial effort would be required for deterministic modeling. Regardless of the feasibility of such complex modeling, understandability of the model would need to be traded in for completeness, also known as *Bonini's paradox* (Dutton & Starbuck, 1971). Also, such complex models would be hard to verify (Roberts & Pashler, 2000) and might generalize poorly due to the large number of free parameters (Vandekerckhove, Matzke, & Wagenmakers, 2015). Alternatively, leaner models would need to emulate information search in the face of epistemic uncertainty. A number of methods have been used to reduce epistemic uncertainty in modeling individual behavior in complex tasks, such as pre-test scores as predictors (Rehling et al., 2004), model-tracing (Fu et al., 2006), inserting physiological data on user's workload into the model (Putze, Schultz, & Propper, 2015), and dynamic adjustment of parameters with pre-computed lookup tables (Fisher, Walsh, Blaha, Gunzelmann, & Veksler, 2016).

(b) When providing assistance, assumptions about the pilot's cognitive state need to be made. Behavioral (including physiological) measures of the pilot's functional states can include contextual information and help in narrowing down the problem space in terms of his or her cognitive state. For a cognitively plausible model of assistance, the model should avoid to make guesses in face of epistemic uncertainty and consult additional

sources of information on the pilot's state instead (Fox & Ülkümen, 2011). Similarly, model-based assistance should follow the path with the greatest prospect of success and not engage in exploration behavior in situations characterized by aleatory uncertainty.

### 1.3. Cognitive assistance in aviation

Increasing automation complexity and task dynamics require pilots to both keep track of the aircraft's state and changes in the situation. While following up with automation modes and transitions puts high demands on pilots' memory, dynamic situations ask a lot of attentional resources to perceive and process incoming alerts and messages on the flight deck. Inattentive deafness has been shown to cause performance drops in the cockpit (Dehais, Roy, & Scannella, 2019) that can be prevented with the help of cognitive assistance, for example in the form of verbal reminders (Estes et al., 2016). Causes and consequences of overheard messages for individual pilots' performance need to be considered to identify what is the right information to be provided and when is the right time to provide it for cognitive assistance in operations.

Causes can be diverse and situation-dependent (e.g., perceptual/attentional factors, see Dehais, Roy, et al., 2019) and are likely too complex for deterministic modeling of single occurrences of missed alerts. Often, alerts are declared as missed when pilots fail to react. Knowing what made a pilot fail to react or what pieces of information he or she was unable to process gives diagnostic value and helps to identify adequate means of support. For cognitive assistance in handling flight deck alerts, information about the contents of a message and whether the message was adequately processed by the pilot is a viable alternative to complex models required for deterministic prediction of user states.

Just as intelligent training systems (Fu et al., 2006), cognitive assistance on the flight deck can be designed in reference to models based on a cognitive architecture. Consequences of an overheard or ignored message for pilots' performance can be anticipated with the help of a cognitive pilot model. ACT-R (Anderson et al., 2004) is a comprehensive and scientifically substantiated cognitive architecture that has produced models representing processes, for example, involved in "manual" flight control of single engine aircraft (Somers & West, 2013), visual attention allocation in a glass cockpit (Byrne et al., 2004), and the use of and skill acquisition for the flight management system (Schoppek & Boehm-Davis, 2004; Taatgen, Huss, & Anderson, 2008). For model-based assistance, such formal descriptions of flight-related tasks and processes can describe what constitutes normative performance.

### 1.4. Neuroadaptive cognitive modeling

In the present paper, a modeling concept is proposed that is able to explain uncertainty in single instances of missed alerts by representing individual pilots' behavior. In the fashion of Putze et al. (2015), we extend the idea of model-tracing (Fu et al., 2006) by incorporating physiological data. Whereas Putze et al. (2015) integrate physiological data to model architectural differences, that is, occupying cognitive resources with a dummy

model to model workload, the concept in this paper focuses on modeling knowledge differences (Taatgen, 1999) due to unprocessed auditory messages.

Model-tracing based on monitoring pilot interactions with flight deck instruments enables the model to identify when performance deviates from normative behavior. Based on such deviations, the model can make inferences about the pilot's cognitive states. By treating instances of deviating behavior as unexplainable situations due to lack of knowledge, the model consults external sources of information, that is, event-related physiological data of the pilot whose cognitive processes it tries to represent.

Physiological measurements, for example electroencephalogram (EEG), can provide information about cognitive operations. With a passive brain-computer interface (Zander & Kothe, 2011), EEG can be recorded without interfering with the task and data can be processed in (almost) real time. The integration of these data into the model allows for more refined representations of individual pilots. Such a neuroadaptive (Zander, Krol, Birbaumer, & Gramann, 2016) cognitive model would be able to adjust its generic or normative behavior to measurements of a pilot's current cognitive state and to identify individual current needs for assistance.

Physiological measures can be subject to errors that introduce intrinsic or aleatory uncertainty (Kiureghian & Ditlevsen, 2009). Whereas epistemic uncertainty represents defined model boundaries, aleatory uncertainty is hard to identify in single situations where there is no ground truth available. That is, the model is able to identify instances of deviating pilot behavior, but it cannot say which of the physiological data are affected by measurement or classification error and which are not. In model-based cognitive assistance, thoughtful handling of the two types of uncertainty is required (see Fig. 1 for an overview of type of uncertainty introduced by data source).

The objective of this study is to increase the effectiveness of modeling individual pilot behavior in response to flight deck alerts. To this end, model-tracing and EEG recordings are used to reduce uncertainty due to individual differences. Behavioral data from an empirical study on the neurophysiological reaction to auditory signals in simulated flight

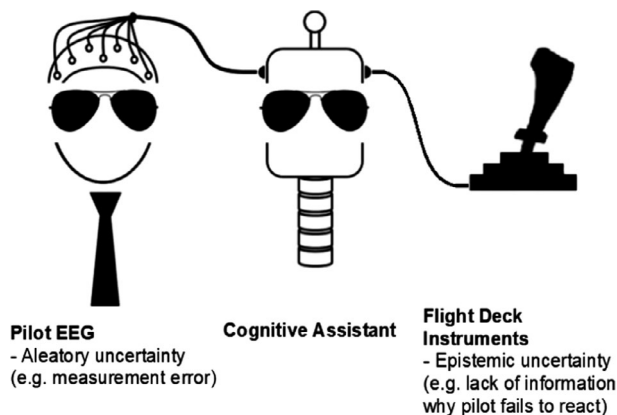


Fig. 1. Sources of uncertainty in neuroadaptive concept.

(Krol et al., 2018) are modeled to demonstrate how the proposed concept can be implemented. Accuracies of a neuroadaptive cognitive pilot model and normative model are compared to quantify the fraction of uncertainty reduced by inserting pilots' EEG data. Epistemic and aleatory uncertainty are quantified and examined regarding their implications for model-based cognitive assistance in flight operations.

## 2. Methods

### 2.1. Empirical study

In all, 21 air crew (one female) who were predominantly military pilots participated in the empirical flight simulator study. Participants had a mean age of 49.08 years ( $SD = 6.08$ ) and an average experience of 3,230 h of flight ( $SD = 2,330.71$ ). All participating air crew had normal or corrected-to-normal vision, and all but two were right-handed. Air crew were seated in a fixed base experimental flight simulator in a single pilot setup that approximated Airbus A320 cockpit design. Participants were asked to perform an 18 min scenario that consisted of 9–14 events resembling flight deck alerts per participant, each preceded by auditory warnings or air traffic control (ATC) messages. The scenario had to be flown by selecting heading and altitude on the auto flight system according to ATC instructions. In addition, participants were asked to manage thrust manually and attend to alerts. Alerts included in the scenario could have low (“amber alert,” e.g., fuel pump failure) or high priority (“red alert,” e.g., engine fire), and ATC messages contained navigation or speed instructions. Speed warnings were issued dynamically whenever participants left a speed range, which resulted in different numbers of acoustic events per participant. For building and simulating the scenario, the open-source flight simulation software “FlightGear 3.4” (<http://home.flightgear.org/>) was used. Essential instrument properties and state changes in the scenario were recorded in log files with a sampling rate of 20 Hz.

Before the flight scenario, participants' EEG was recorded while performing an auditory oddball paradigm (frequent versus rare sounds). A classification algorithm was trained on the EEG data to recognize activity patterns for processing of target (i.e., processed alerts) and standard sounds (missed alerts). The algorithm was tuned to have equal chances for false alarms and misses in case of incorrect response classification. Due to the frequent use of standard compared to rare target sounds in the training paradigm, classifier accuracy needs to be higher than 0.78 to perform significantly better than chance. EEG was recorded during classifier training and scenario with a 32-channel BrainProducts LiveAmp system.

### 2.2. Cognitive modeling

A cognitive model was created to represent individual pilots' behavior using the cognitive architecture ACT-R version 7.14. ACT-R consists of memory, perceptual, and motor



modules that interact with each other by exchanging chunks of information through buffers. The declarative memory module can hold and store information about the task state, whereas procedural memory allows for modeling productions (condition–action–statements) that apply depending on the state of the task or the environment. Perceptual and motor modules allow for modeling of basic sensory processes and enable a model to interact with the environment. When modeling pilot activities, the respective modules can be used to represent storing and updating flight information such as altitude and speed, procedures for how to react in case of alerts, and auditory and visual perception of messages in the cockpit.

For assistance in operations, a cognitive pilot model needs to be flexible, adaptive at runtime, and knowledgeable of the operational context. Not only does it need to know what constitutes optimal or normative performance of a task but also alternative means to meet the objective. In case of deviations from normative performance, it has to be able to adapt its functionality and adjust its representation of the pilot. Finally, the model needs to be able to anticipate the consequences of both normative and alternative performance in a task so it can offer support when needed.

A scenario-specific hierarchical task analysis (HTA; Stanton, 2006) was conducted for identifying seven main tasks of which one routine and six alert specific tasks. Main tasks were then split up iteratively until the lowest level of actions that can be observed in simulator log files. Based on this HTA, an ACT-R cognitive model was created that was able to memorize flight information by reading airspeed and altitude data, decide when to adjust the throttle, process and respond to auditory messages, and check if its own actions match pilot's actual behavior. This model will be referred to as the "normative" model.

A Python implementation of the extended version of ACT-CV (Halbrügge, 2013) was used to create an interface between the cognitive architecture and FlightGear log files. While ACT-CV allows real-time interaction between cognitive models and the outer world, the possibility of fast-time simulation was considered more important in the given scenario. Therefore, ACT-R did not interface with FlightGear directly (see Somers & West, 2013), but through recordings of individual participants' performance. The graphical interface of the flight simulator was presented textually in the visicon, that is, ACT-R's visual representation of the environment. As an accurate representation of pilots' visual behavior was beyond the scope and focus of this study, different parameters (e.g., airspeed, altitude, etc.) were presented at predefined locations independent of Airbus cockpit design.

Parameter changes linked to events (e.g., engine1-on-fire from "0" to "1") triggered sounds in the audicon, that is, ACT-R's acoustic scene, so messages from the cockpit were presented in the same modalities as in the empirical study. A processed EEG data were displayed as event-related Boolean variable ("1" for alerts processed as target sound, "0" for standard sounds). Alert procedures in the form of checklists as well as contents of ATC messages in the controller-pilot datalink communications could not be communicated through FlightGear. As a workaround, an extra buffer was added that gives the model access to information not displayed in the visicon. Checklist and ATC information was therefore defined as chunks in declarative memory containing information, for example, about the event name (e.g., "engine-fire"), the current action to perform

in terms of button or dial (e.g., “engine-cutoff”) and the value to be set (e.g., “true”), and the next step on the checklist (e.g., “fire-bottle-discharge,” “true”). Modeling the reading- or listening-intensive processes of following checklists or processing ATC information is beyond the scope of this model. The additional retrieval buffer allowed the model to access the information without occupying other buffers. The Python module included in ACT-R version 7.14 was used for interfacing with ACT-CV, running the model, and storing results.

2.2.1. Normative model

For the routine task (see Fig. 2), the model monitors variables of airspeed and altitude that were shown in the simulator’s primary flight display. In alternating order, the model updates its internal representation of the flight information that is stored as declarative

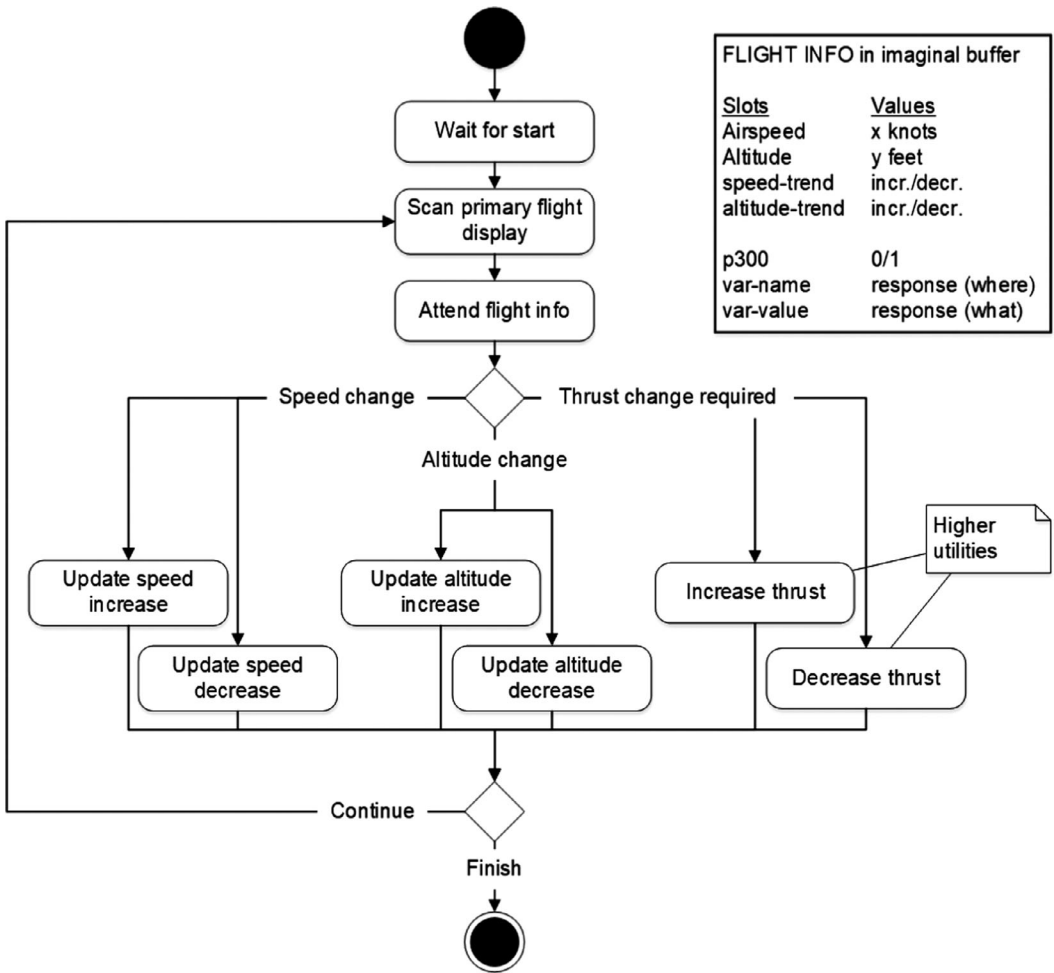


Fig. 2. Routine loop in both models.



knowledge in the imaginal buffer. When updating its information on altitude and speed, the model also computes trends for speed and altitude in “decreasing” or “increasing” by comparing the most recent values in the imaginal buffer to the new value in the visual buffer. If the model notices that airspeed deviates from target speed and altitude, and if the speed-trend suggests this deviation will grow larger, the model adjusts the thrust accordingly. Utilities of thrust manipulation productions are set to an increased value to ensure they are selected over passive monitoring productions. If no adjustment of thrust is required, the model returns to monitoring speed or altitude after updating its flight information.

In case of auditory signals, the normative model temporarily exits this routine loop, processes the sound, and shifts visual attention to read the corresponding warning message or event-name. In case of ATC messages, the model retrieves the chunk corresponding to the event-name, processes navigational instructions in the chunk (e.g., “heading-select,” “180”), and stores them in the imaginal buffer. If the model hears an alert, it retrieves a checklist corresponding to the specific alert and puts the action required from the pilot in the imaginal buffer. For all acoustic events, the normative model assumes that pilots correctly process the information and will respond adequately. After each event, it checks the log data for the required pilot response to evaluate if its assumption is correct. Situations where pilots do not respond adequately are treated as epistemic uncertainty and marked as cases when some sort of assistance should be provided or communication from the pilot to the model are needed for more accurate state classification.

### 2.2.2. Neuroadaptive model

The neuroadaptive model forms an extension of the normative model. It follows the same courses of action for routine tasks and acoustic events that were followed by an adequate pilot response. In cases of epistemic uncertainty when no adequate response is observed, the neuroadaptive model emulates information search behavior and consults the EEG data to check if the pilot had paid attention to the sound (see Fig. 3). If EEG data show the pilot has processed the alert or message like a standard instead of a target stimulus, the neuroadaptive model updates its description of the situation to a missed alert. The model considers these cases as situations that require verbal reminders of the alert or message. Situations where no adequate response was observed but EEG data show the preceding sound was processed correctly are treated as epistemic uncertainty. For these situations, the model knows that assistance of some form other than a verbal reminder is needed.

### 2.3. Analysis

After each acoustic event, the first reaction to the auditory events was evaluated, for example adjusting the selected altitude in response to ATC messages. Adequacy and timeliness of responses were scored according to criteria assessed in the HTA with subject matter experts. While adequacy of responses depended on the type of alert or contents of ATC messages, the time limit for initiating a first reaction to an alert was set to 25 s for all events. For example, if an ATC message requested a flight level change to

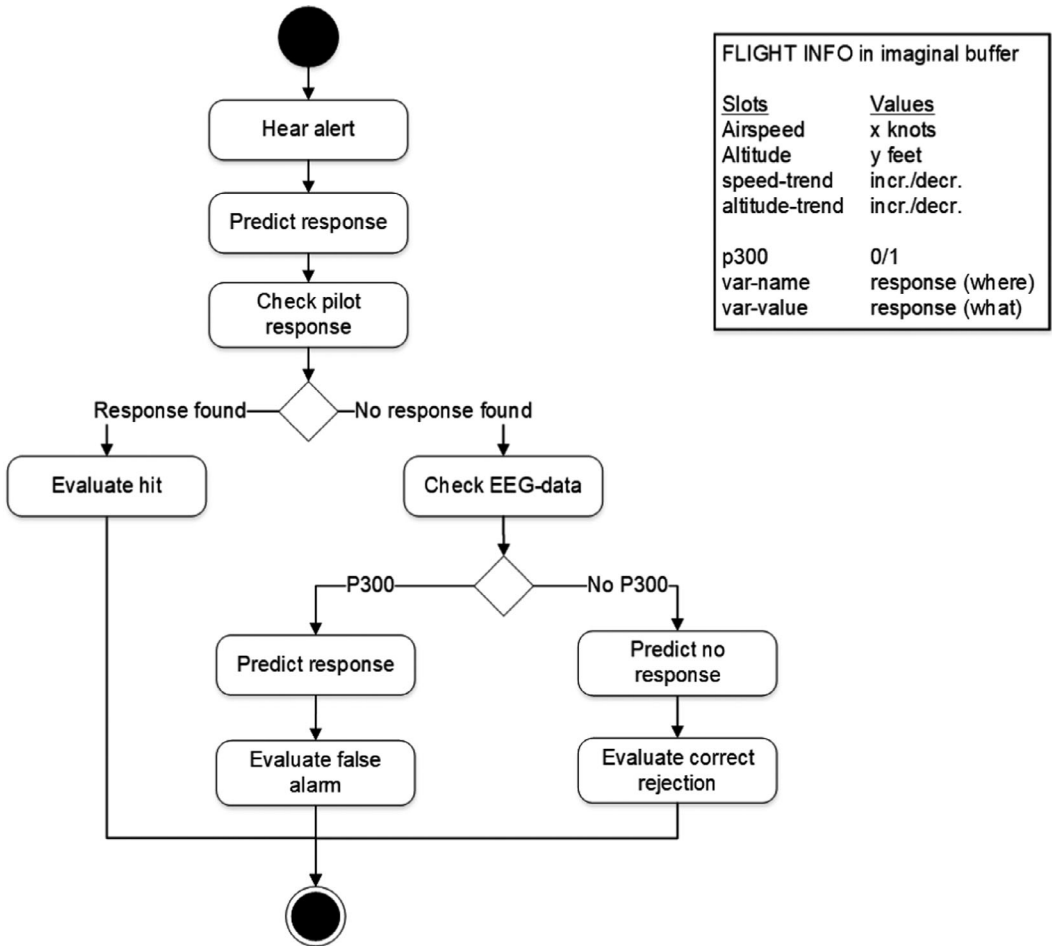


Fig. 3. Alert procedure for neuroadaptive model.

300, entering an altitude-select of 300 in the flight control unit within a time window of 25 s was scored as good performance; all other responses such as entering an altitude-select of 280 or entering the correct altitude-select after 25 s were classified as missed ATC message. In both models, epistemic uncertainty was scored as incorrect description of pilot behavior. Both the normative and the neuroadaptive model could correctly describe situations with adequate pilot reactions to acoustic events; in addition, the neuroadaptive model was able to classify lacking responses as correct descriptions when EEG data showed that the sound was not processed adequately.

Correctly described responses are scored with 1, incorrect response descriptions with 0. For each participant, both models divide the sum of correct descriptions by the total number of alerts and ATC messages to quantify model accuracy. For both models, mean accuracy is computed across pilots. As the number of auditory events was not the same

for all participants due to ATC speed messages, median and interquartile range had to be used as measures of central tendency and dispersion. A Wilcoxon signed rank test for pairwise comparisons was used to quantify the added value of EEG data for the neuroadaptive model.

Aleatory uncertainty in the neuroadaptive model is equal to one minus EEG classifier accuracy. As the data give no information about which situations are concerned by classifier inaccuracies, the model emulates exploitation behavior by accepting aleatory uncertainty of the EEG classifier and relying on its classification given that its accuracy is above chance. Added value of neuroadaptivity to the normative model was quantified by subtracting normative from neuroadaptive model accuracy. By multiplying added value with EEG classifier accuracy, a mean accuracy of the neuroadaptive model corrected for aleatory uncertainty was computed. Chi-square tests were used to test for relationships between event types and model classification. To this end, events were grouped into categories “ATC messages,” “Fuel pump failure,” “Fire alert,” and “Speed warnings.” For the subset of missed alerts, a chi-square test was performed to test for a possible relationship between the type of responses failure (i.e., incorrect response or no response at all) and EEG classifier output (alert processed or not). Where required, graphs are shown to provide information about possible interaction effects in the data that could not be explored by the non-parametric tests used in the analysis. Starting from a level of significance of 0.05, all tests are conducted at an alpha corrected for the number of tests using Bonferroni’s method of  $\alpha' = 0.05/3 \approx 0.02$ .

### 3. Results

In total, behavior descriptions for 226 events were generated by each model for all pilots with an average of 10.8 ( $SD = 0.9$ ) per pilot. The normative model correctly

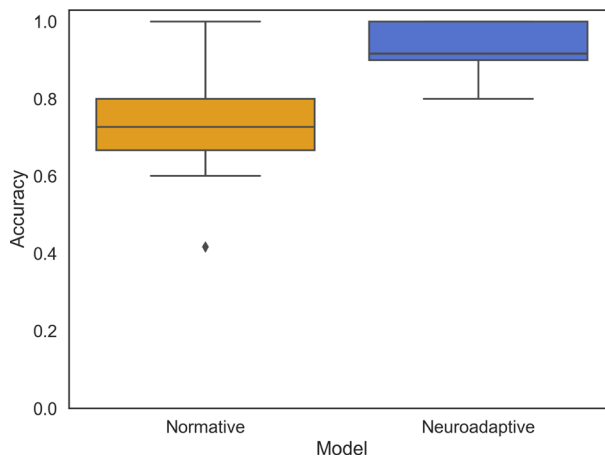


Fig. 4. Median accuracy per model.

described participant's behavior for 165 of these events ( $Acc_{Norm.} = 0.73$ ) with a median model accuracy of  $MDN_{Norm.} = 0.73$  ( $IQR = 0.80-0.67$ ; Fig. 4). Thus, the total amount of uncertainty treated as epistemic is  $u_{Epistemic} = 0.27$ .

The neuroadaptive cognitive model generated correct descriptions in 213 of 226 cases ( $Acc_{Neuro.} = 0.94$ ) with a median accuracy of  $Mdn_{Neuro.} = 0.92$  ( $IQR = 1.0-0.9$ ; Fig. 4). The uncertainty treated as epistemic is therefore  $u_{Epistemic} = 0.06$ .

The signed rank test showed that neuroadaptive model accuracy is significantly higher compared to the normative model ( $z = -4.01, p < .001$ ). Added value of the EEG data is 0.21. Correcting the added value for the EEG classifier accuracy of 0.86 resulted in a corrected accuracy of the neuroadaptive model of  $Acc_{Neuro.}' = 0.91$  and aleatory uncertainty of  $u_{Aleatory} = 0.03$ . Model accuracies per participant and model are shown in Fig. 5.

According to the chi-square test, there was no significant effect of the type of event on the mean accuracy across participants and model types ( $\chi^2(3, N = 226) = 0.25, p = .97$ ). Fig. 6 shows the fractions of correct model classifications per event type and model.

The relationship between the type of failed response (incorrect or no response) and the EEG classifier output (processed/not processed) was marginally significant ( $\chi^2(1, N = 61) = 1.04, p = .05$ ), but not significant according to the Bonferroni-corrected  $\alpha'$ . This suggests that the EEG data are not indicative of participants responding incorrectly or not at all to an alert or message. The distribution of alert types per classifier output is shown in Fig. 7.

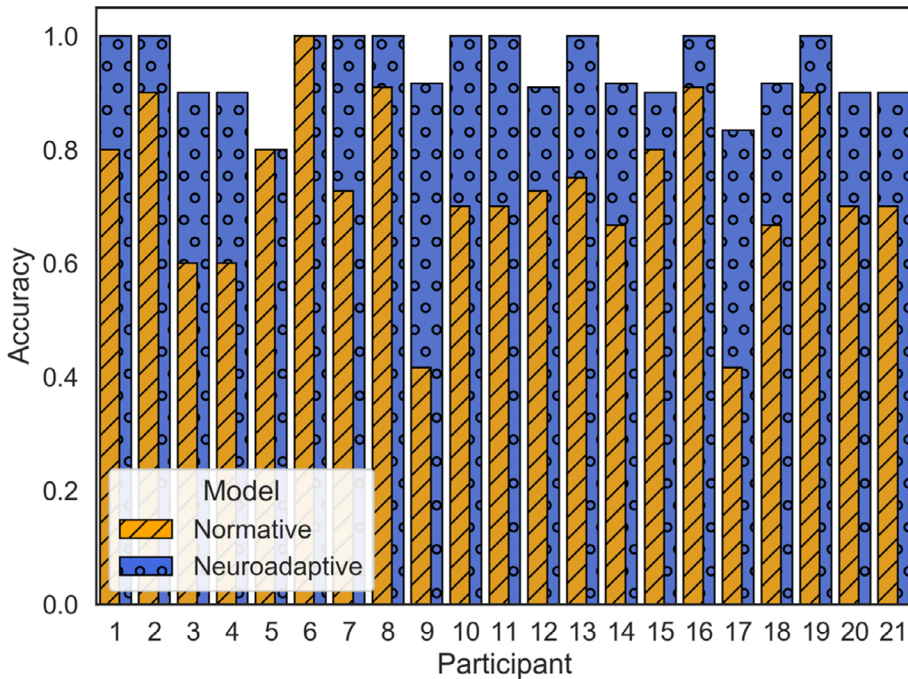


Fig. 5. Mean accuracy per participant and model.

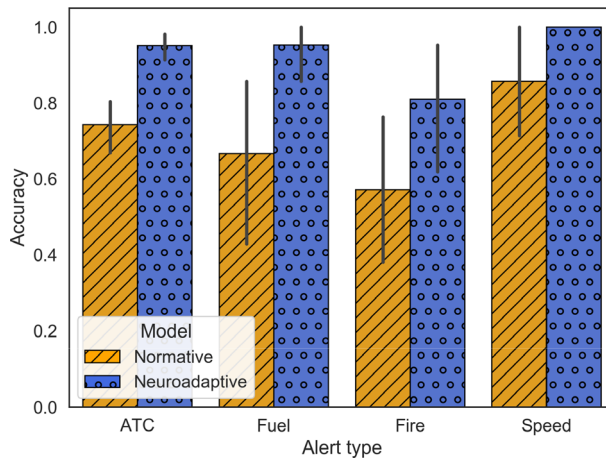


Fig. 6. Mean accuracy per alert type and model. For testing for a relationship between alert types and the model accuracies, events were grouped into the following alert types: ATC messages = “ATC,” fuel pump alert = “Fuel,” engine fire alert = “Fire,” speed warnings = “Speed.”

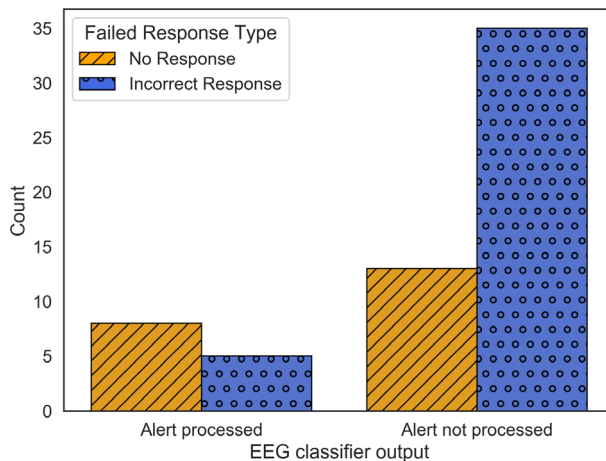


Fig. 7. Frequencies of false response types by classifier output from the subset of missed alerts.

#### 4. Discussion

The presented concept and its application demonstrate how pilot performance can be modeled in spite of individual differences using model-tracing and physiological data. The distinction between aleatory and epistemic uncertainty (Kiureghian & Ditlevsen, 2009) and their quantification was decisive for the neuroadaptive model’s design and implementation. Data show how model accuracy can be significantly increased by connecting model-tracing and EEG data in line. The specification of remaining fractions of epistemic and aleatory uncertainty provides starting points for further improvement of the concept.

Whereas flight deck instrument interactions can be observed directly, unprocessed alerts can only be detected by behavioral or physiological symptoms. Due to aleatory uncertainty introduced by the EEG classifier, model-tracing with instrument and EEG data had to be connected in line to maximize the effectiveness in reducing epistemic uncertainty. Compared to other studies using EEG data to model effects of individual differences, the integration of EEG data was quite straightforward for the neuroadaptive model and did not require a dual model approach (Putze et al., 2015). Model-tracing based on the log files proved effective in detecting deviations from normative behavior due to an increased density of acoustic events in the scenario. Real flight, however, contains long periods of monitoring instruments without direct input required. Deriving mental states based on model-tracing (Fu et al., 2006) in such highly automated or autonomous environments could therefore require other pilot behavior data sources, for example unobtrusive monitoring of neurophysiological activity, speech, or gaze. Cognitive models are well suited for the interpretation of such data by linking physiological phenomena to context.

Apart from measurement and classification errors, the neuroadaptive model was able to explain ~78% of the normative model's epistemic uncertainty, leaving a total of 6% of cases when the model does not know what made participants fail to react adequately. These data suggest that cognitive assistance in the form of verbal reminders would suffice to help with performance recovery in all other situations lacking responses from participants.

Normative model accuracy represents the effects of individual differences on performance given the scenario. By design, the neuroadaptive model improves on the normative model; the significance of improvement with the EEG data is moderated by the effect of individual differences. Nonetheless, increased accuracy of the neuroadaptive model shows how epistemic uncertainty can be reduced with the help of physiological data. For an empirical evaluation of the concept, a comparison with alternative designs for model-based assistance is required. For example, a Wizard-of-Oz setup with a human co-pilot interpreting pilot behavior could be used to compare the neuroadaptive model's accuracy in classification of pilots' cognitive states to that of human co-pilots.

The neuroadaptive model tracks pilots' perception of auditory events. The fact that a piece of information has been perceived and processed by a pilot does not mean that it has been understood. Measures of pilots' situation assessment and awareness (Endsley, 1995) may help to reduce epistemic uncertainty about why a pilot may fail to respond adequately. Physiological symptoms of cognitive conflict, for example in the anterior cingulate cortex (Anderson, Anderson, Ferris, Fincham, & Jung, 2009), could be used to identify when information that was perceived could not be comprehended by the pilot. In addition, the combination of such measures with behavioral data might not only help to identify such conflicts, but it can also provide diagnostic information on what interfering goals may have caused the conflict. Based on such information, cognitive assistance could help pilots to build up or activate more useful mental representations in situations of cognitive conflict, for example in response to automation surprises (Sarter, Woods, & Billings, 1997).



Mean accuracy of the neuroadaptive model corrected for aleatory uncertainty is 91%. Aleatory uncertainty may be reduced with other independent physiological measures, for example eye tracking. EEG classification could be supported with corresponding gaze data by connecting both methods in line. For example, when the EEG data show that a pilot has processed an alert, saccades to the warning display after the alert can reduce the uncertainty by eye tracking classification accuracy. This linear approach can enable cognitive assistance to make Bayesian inferences about individual pilots' cognitive states based on conditional probabilities of cognitive processes or actions given the physiological and other behavioral data.

Further research is required on how to model individual differences in behavior with the help of behavioral and physiological measures. As the objective of our model was to represent the pilot's current state of knowledge about a given situation, individual differences were modeled as knowledge differences by withholding information depending on EEG data. More comprehensive approaches that, for example, integrate measures of pilots' resource availability (e.g., Dehais, Rida, et al., 2019) might be able to anticipate missed alerts and to model pilots' differences in knowledge about a situation as a function of their architectural differences (Taatgen, 1999). Albeit being more complex, such modeling of architectural differences could provide diagnostic information and thereby explain why certain messages were not processed correctly in a complex task.

No effect of the type of event on the models' accuracies was found. Fig. 6 does not show signs of an interaction effect between the types of events and models, which suggests that the different event types neither affected the chance of responding correctly (i.e., the normative model results) nor the EEG classifier output (i.e., as integrated by the neuroadaptive model). This is unexpected as the events differ substantially in their salience and associated criticality. Due to the study's limitations, for example regarding the simulator setting, density of events in the scenario, and potential observer effects, this finding may not generalize well to other experimental setups, not to mention real-world scenarios.

The integration of EEG data in the cognitive model could not provide clear information about whether participants responded incorrectly or whether they do not respond at all, which suggests that the EEG classifier was sensitive to missed and incorrectly processed alerts alike. However, results of the test for a relationship between type of failed response and EEG classifier output approached significance and the difference between distributions might become clearer with larger sample sizes. Fig. 7 indicates a larger fraction of incorrect responses for alerts that were classified as incorrectly processed based on the EEG data, while failed alert types are more evenly distributed for correctly processed sounds. In reality, the distinction between missed and incorrectly processed alerts might be more gradual than suggested by the analysis in this study. In-depth analyses of neurophysiological measures can give insights into what processing stages a stimulus or piece of information passes through on a single-trial basis (e.g., perceptual, dorsal or ventral auditory pathway, etc.). Alternatively, the imbalance of incorrect and no responses in missed alerts could be explained by strategies of coping with uncertainty: If uncertainty about the contents of incorrectly processed messages was considered being aleatory,

participants may have been more likely to take a gamble and provide some response rather than none (Fox & Ülkümen, 2011). Further research is required to understand and model in what cases pilots respond incorrectly or not at all after incorrectly processed alerts.

As a first step toward pilot action anticipation for cognitive assistance, the modeling concept took inspiration from fundamental abilities of (a) sharing representations and (b) predicting actions described by Sebanz et al. (2006). (a) Following up with perceptual processes of the operator can allow for inferences about the pilot's representation of the situation. Similar to anticipation in human–human interactions, being receptive to emotional or overt physiological reactions can help in validating assumed representations. (b) Recognizing and understanding an individual's actions at a given moment is a necessary but not a sufficient condition for successful cooperation and joint action. Rather, the ability to predict outcomes of an individual's actions and knowing what actions will likely follow enables the coordination of efforts. For example, tracing the perceived amount of uncertainty or confidence in operators' representation of a situation can help anticipating the timing of their actions (e.g., Scharfe & Russwinkel, 2019).

The study described in the paper has given an example of how model-based assistance in human machine interaction can provide machines with an implicit feedback loop that allows for checking if the information they provide is perceived by the operator. Building on cognitive models that integrate physiological data about operators' states, future research should aim at tracing operators understanding of a situation in the form of mental representations in operations. Ideally, this will enable machines to form a more refined model of their users, to share their representations, and to anticipate their behavior in much the same way that humans learn to interact and cooperate with humans.

## Acknowledgments

The authors would like to thank Matthias Buderath, Daniel Dreyer, Lena Andreessen, Dan Bothell, BrainProducts GmbH, and Inge Wetzels for their support.

## References

- Anderson, J. R., Anderson, J. F., Ferris, J. L., Fincham, J. M., & Jung, K.-J. (2009). Lateral inferior prefrontal cortex and anterior cingulate cortex are engaged at different stages in the solution of insight problems. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10799–10804. <https://doi.org/10.1073/pnas.0903953106>
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. <https://doi.org/10.1037/0033-295X.111.4.1036>
- Byrne, M. D., Kirlik, A., Fleetwood, M. D., Huss, D. G., Kosorukoff, A., Lin, R.-S., & Fick, C. S. (2004). A closed-loop, ACT-R approach to modeling approach and landing with and without synthetic vision system (SVS) technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(17), 2111–2115. <https://doi.org/10.1177/154193120404801707>

- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*(19), 1905–1910. <https://doi.org/10.1016/j.cub.2006.07.065>
- Dehais, F., Rida, I., Roy, R. N., Iversen, J., Mullen, T., & Callan, D. (2019). A pBCI to predict attentional error before it happens in real flight conditions. In M.P. Fanti, M. Zhou, D.B. Goldgof, R. Roberts (Eds.), *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 4155–4160). IEEE. <https://doi.org/10.1109/SMC.2019.8914010>
- Dehais, F., Roy, R. N., & Scannella, S. (2019). Inattentive deafness to auditory alarms: Inter-individual differences, electrophysiological signature and single trial classification. *Behavioural Brain Research*, *360*, 51–59. <https://doi.org/10.1016/j.bbr.2018.11.045>
- Dutton, J. M., & Starbuck, W. H. (1971). *Computer simulation of human behavior*. New York: Wiley.
- Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65–84. <https://doi.org/10.1518/001872095779049499>
- Estes, S., Burns, K., Helleberg, J., Long, K., Stein, J., & Pollack, M. (2016). Digital copilot: Cognitive assistance for pilots. In F. Stein (Ed.), *Proceedings of the AAAI fall symposium on cognitive assistance in government and public sector applications (Technical Report FS-16-02)* (pp. 141–144). Palo Alto, CA: The AAAI Press.
- Fisher, C. R., Walsh, M. M., Blaha, L. M., Gunzelmann, G., & Veksler, B. (2016). Efficient parameter estimation of cognitive models for real-time performance monitoring and adaptive interfaces. In D. Reitter & F. E. Ritter (Eds.), *Proceedings of the 14th International Conference on cognitive modeling* (pp. 113–118). Pennsylvania, PA: Penn State.
- Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.), *Perspectives on thinking, judging, and decision making* (pp. 21–35). Oslo: Universitetsforlaget.
- Fu, W.-T., Bothell, D., Douglass, S., Haimson, C., Sohn, M.-H., & Anderson, J. (2006). Toward a real-time model-based training system. *Interacting with Computers*, *18*(6), 1215–1241. <https://doi.org/10.1016/j.intcom.2006.07.011>
- Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to “mirror” and “canonical” neurones in the human brain: An fMRI study. *NeuroImage*, *18*(4), 928–937. [https://doi.org/10.1016/s1053-8119\(03\)00042-9](https://doi.org/10.1016/s1053-8119(03)00042-9)
- Halbrügge, M. (2013). ACT-CV – Bridging the gap between cognitive models and the outer world. In E. Brandenburg (Ed.), *Grundlagen und Anwendungen der Mensch-Maschine-Interaktion: 10. Berliner Werkstatt Mensch-Maschine-Systeme, 10.–12. Oktober 2013. Berlin = Foundations and applications of human machine interaction* (pp. 205–210). Berlin: TU Berlin.
- Kiureghian, A. D., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, *31*(2), 105–112. <https://doi.org/10.1016/j.strusafe.2008.06.020>
- Krol, L. R., Klaproth, O. W., Vernaleken, C., Wetzel, I., Gaertner, J., Russwinkel, N., & Zander, T. O. (2018). Towards a neuroadaptive cockpit: First results. Third International Mobile Brain/Body Imaging Conference, Berlin, Germany.
- Putze, F., Schultz, T., & Propper, R. (2015). Dummy model based workload modelling. In S. Kwong (Ed.), *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 935–940). Hong Kong: IEEE. <https://doi.org/10.1109/SMC.2015.171>
- Rehling, J., Lovett, M., Lebiere, C., Reder, L., & Demiral, B. (2004). Modeling complex tasks: An individual difference approach. In K. Forbus, D. Gentner, & T. Reiger (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 26, pp. 1137–1142). Chicago, IL: Psychology Press.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367. <https://doi.org/10.1037/0033-295x.107.2.358>
- Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926–1943). New York: Wiley.

- Scharfe, M., & Russwinkel, N. (2019). Towards a cognitive model of the takeover in highly automated driving for the improvement of human machine interaction. In T. C. Stewart (Chair), *Proceedings of the 17th International Conference on cognitive modelling*. Waterloo, Canada: University of Waterloo.
- Schoppek, W., & Boehm-Davis, D. A. (2004). Opportunities and challenges of modeling user behavior in complex real world tasks. *MMI Interaktiv*, 7, 47–60.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76. <https://doi.org/10.1016/j.tics.2005.12.009>
- Somers, S., & West, R. (2013). Steering control in a flight Simulator using ACT-R. In R. L. West & T. C. Stewart (Eds.), *Proceedings of the 12th International Conference on cognitive modeling* (pp. 233–238). Ottawa: Carleton University.
- Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions. *Applied Ergonomics*, 37(1), 55–79. <https://doi.org/10.1016/j.apergo.2005.06.003>
- Taatgen, N. A. (1999). Cognitief Modelleren: Een nieuwe kijk op individuele verschillen. *Nederlands Tijdschrift Voor De Psychologie*, 54(4), 167–176.
- Taatgen, N. A., Huss, D., & Anderson, J. R. (2008). The acquisition of robust and flexible cognitive skills. *Journal of Experimental Psychology: General*, 137(3), 548–565. <https://doi.org/10.1037/0096-3445.137.3.548>
- Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of computational and mathematical psychology* (Vol. 1, pp. 300–319). Oxford, UK: Oxford University Press.
- Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., & Wahn, B. (2016). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology*, 7, 2039. <https://doi.org/10.3389/fpsyg.2016.02039>
- Zander, T. O., & Kothe, C. (2011). Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general. *Neural Engineering*, 8(2), 25005. <https://doi.org/10.1088/1741-2560/8/2/025005>
- Zander, T. O., Krol, L. R., Birbaumer, N. P., & Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proceedings of the National Academy of Sciences of the United States of America*, 113(52), 14898–14903. <https://doi.org/10.1073/pnas.1605155114>
- Zhang, Z., Russwinkel, N., & Prezenski, S. (2018). Modeling individual strategies in dynamic decision-making with ACT-R: A task toward decision-making assistance in hCI. *Procedia Computer Science*, 145, 668–674. <https://doi.org/10.1016/j.procs.2018.11.064>